2011

TR-2011007: Randomized and Derandomized Matrix Computations

Victor Y. Pan

Guoliang Qian

Ai-Long Zheng

Randomized and Derandomized Matrix Computations *

Victor Y. Pan[1,2],[a], Guoliang Qian[2],[b], and Ai-Long Zheng[2],[c]

[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
[c] azheng-1999@yahoo.com

Abstract

We propose new techniques and algorithms that advance the known methods for a number of fundamental problems of matrix computations. These problems includes approximation of leading and trailing singular spaces of a matrix with extensions to derandomized approximation by low-rank matrices and by structured matrices, support for numerically safe Gaussian elimination with no pivoting, and devising effective preconditioners that cover the general class of matrices having a small numerical nullity or a small numerical rank. Our technical novelties include randomized additive preconditioning and augmentation for general and structured matrices, derandomization, and dual extension of the Sherman–Morrison–Woodbury formula. Our extensive tests demonstrate effectiveness of the proposed algorithms.

Key Words: Low-rank matrices, Approximation, Linear systems of equations, Randomized preconditioning, Dearndomization, Numerical rank

1 Introduction

1.1 Derandomized approximation by low-rank matrices

Approximate matrix decompositions based on approximation of matrices A having small numerical ranks q by low-rank matrices is a thriving research area with numerous important applications and extensions [HMT11]. To the wealth of these extensions we can add approximation by matrices with displacement structure (see Section 5.2) and tensor computations [T00], [MMD08], [OT09].

The approximation is actually needed just for the range of A, and the most popular algorithms employ multiplication of an $m \times n$ input matrix A by $q+p$ random vectors where $p \ll q \ll \min\{m, n\}$; then with a high probability the desired approximation can be readily recovered [HMT11].

*Supported by NSF Grant CCF-1116736 and PSC CUNY Awards 62230–0040 and 63153–0041. 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, and the 3nd International Conference on Matrix Methods in Mathematics and Applications (MMMA 2011) in Moscow, Russia, June 22-25, 2011

In this paper we present an alternative algorithm (see Section 5.2) where the desired approximation is given by the range of the matrix C_-U_- where $C_- = A - AU_-H^{-1}V_-^TA$ for $H = I_q + V_-^TAU_-$, U_- and V_- are random matrices of the sizes $m \times q$ and $n \times q$, respectively, and M^T denotes the transpose of a matrix M. To our advantage, we get rid of the extraneous parameter p and derandomize the computations: by scaling matrices U and V we ensure that the matrix H is diagonally dominant. One of our goals is to advance this algorithm further by combining it with various technically distinct algorithms of [HMT11], [GTZ97], [GT01], [GOS08] and [HMT11].

1.2 Numerically safe Gaussian elimination with no pivoting

To motivate our further study we recall that the condition number $\kappa(A)$ of a matrix A of a rank ρ is the ratio $\sigma_1(A)/\sigma_\rho(A)$ where $\sigma_j(A)$ is the jth largest singular value of the matrix A for $j = 1, \dots, \rho$. If $\kappa(A)$ is large (in context), then the matrix A is called ill conditioned and lies near a rank deficient matrix; its inversion and solving linear systems with such a matrix are readily corrupted if they are performed numerically with rounding errors by using single or double precision. Otherwise A is well conditioned, and if it also has full rank, then numerical computations are safe for it unless they involve ill conditioned auxiliary matrices.

In particular pivoting, that is row or column interchange is applied to avoid dealing with such auxiliary matrices. Gaussian elimination with no pivoting (hereafter we refer to it as GENP) can easily fail in numerical computations with rounding errors, except for some special classes of input matrices, in particular diagonally dominant or positive definite matrices. For such classes GENP and its pivoting-free variations outperform Gaussian elimination with pivoting (cf. [GL96, page 119]). We dramatically expand these classes by proving in Section 5.1 that pre- as well as post-multiplication of a well conditioned coefficient matrix by a Gaussian random square matrix supports safe numerical performance of GENP as well as block Gaussian elimination. In our tests this recipe consistently worked even when we applied circulant multipliers and filled their fisrt columns with the values one and -1, thus limiting randomization to the choice of the signs $+$ or $-$.

1.3 Randomized and derandomized preconditioning of matrices with a small numerical nullity or a small numerical rank

Can we extend the latter advance by applying randomized multipliers M and N to precondition an ill conditioned input matrix A, that is to yield a better conditioned matrix product? This does not work because $\kappa(A) \le \kappa(MAN)\kappa(M)\,\mathrm{cond}(N)$ and because random matrices are expected to be well conditioned [D88], [E88], [CD05], [SST06], [ST02]. Approximate inverses $X \approx A^{-1}$ are popular as multipliers, but they are readily computed only for some special although important classes of matrices.

Traditionally the transition $A \Longrightarrow B$ is called *preconditioning* wherever it simplifies the solution of a linear system $A\mathbf{y} = \mathbf{b}$, e.g., where B is the product MA, AN or MAN and the matrix $I - B$ has a small numerical rank for I denoting the identity matrix. Indeed in this case some most popular iterative algorithms such as CG and GMRES algorithms converge to the solution of a linear system $B\mathbf{x} = \mathbf{f}$ very fast even if $\mathrm{cond}(B)$ is not small. Here and hereafter we use the acronym "CG" for "Conjugate Gradient".

Additive preprocessing $A \Longrightarrow C = A + B$ yields $C = I$ for $B = I - A$, but this observation is not easy to utilize for solving a linear system $A\mathbf{y} = \mathbf{b}$. Assume, however, that an $n \times n$ nonsingular input matrix A scaled so that $||A||_2 \approx 1$ has a numerical nullity at most r, that is has at most r singular values that are much smaller than one. Further assume that U and V are $n \times r$ standard Gaussian random matrices and write $C = A + UV^T$. Then we prove (cf. Corollary 4.3) that with a high probability $\kappa(C)$ has order $\sigma_1(A)/\sigma_{n-r}(A)$ versus $\kappa(A) = \sigma_1(A)/\sigma_n(A)$.

We prove this upper bound on $\kappa(C)$ for any matrix A; furthermore with a small additional work we can derandomize our preprocessing to produce $n \times r$ matrices U_1 and V_1 such that $\kappa(C_1) \approx \sigma_1(A)/\sigma_n(A)$ for $C_1 = A + U_1V_1^T$; $\kappa(C_1) = \sigma_1(A)/\sigma_n(A)$ if A has nullity r (see Section 5.3).

The ratio $\sigma_1(A)/\sigma_{n-r}(A)$ is not large for matrices A with positive numerical nullity at most r, which is a large and important subclass in the class of ill conditioned matrices (cf. [CDG03] and

our Remarks 2.1 and 4.4); therefore the matrices C_1 and C associated with such matrices A are well conditioned with probability one and near one, respectively.

Such matrices can be more readily inverted than an ill conditioned matrix A; consequently their inverses can be used as multiplicative preconditioners for it. Indeed the Sherman–Morrison–Woodbury formula (hereafter referred to as *SMW formula*) implies that the matrices $I-AC^{-1}$, $I-C^{-1}A$, $I-AC_1^{-1}$, and $I-C_1^{-1}A$ have ranks at most r.

Likewise the matrices C_-^{-1} for $C_-=A-AU_-H^{-1}V_-^TA$, $H=I_q+V_-^TAU_-$, and random scaled matrices U_- and V_- (cf. Section 1.1) can be used as effective preconditioners for nonsingular matrices A having small numerical ranks q. This is because $C_-^{-1}=A^{-1}+U_-V_-^T$, and so the matrices $I-C_-^{-1}A$ and $I-AC_-^{-1}$ have ranks at most q, and because the matrices C_- are expected to be well conditioned are therefore more readily invertible than the matrices A.

Thus our study extends the *analysis of conditioning* in [D88], [E88], [CD05], [SST06], [ST02] to *randomized preconditioning* of matrices having a small numerical nullity or a small numerical rank.

In an alternative application of our additive preprocessing (see Section 5.4), we compute the matrix $W=\begin{pmatrix}V^TC^{-1}\\X\end{pmatrix}A$ for a random scaled $q\times n$ matrix X and deduce that the $r\times r$ leading block of W strongly dominates the subdiagonal block underneath; block Gaussian elimination readily eliminates the latter block and thus reduces the original ill conditioned linear system $A\mathbf{y}=\mathbf{b}$ to two well conditioned systems of smaller sizes. The reduction requires extended precision but uses only $O(n^2r)$ flops versus $\frac{2}{3}n^3+O(n^2)$ in Gaussian elimination. Here and hereafter "flop" stands for "arithmetic operation".

Yet another (more involved) alternative is to compute the solution $\mathbf{y}=A^{-1}\mathbf{b}$ to a linear system $A\mathbf{y}=\mathbf{b}$ by substituting the SMW expression for A^{-1} via C^{-1} (see [PGMQ], [Pa], and our Section 5.7). In the case of ill conditioned matrices A having a small numerical nullity the resulting algorithm solves a linear system $A\mathbf{y}=\mathbf{b}$ by using a nearly optimal number of bitwise (Boolean) operations.

1.4 Approximation of trailing and leading singular spaces

Our algorithms for approximation by low-rank matrices in Section 1.1 and for block triangular reduction above rely on approximation of trailing and leading singular spaces of a matrix, which is a problem of independent importance.

We employ the following basic properties (cf. Theorem 5.2).

(a) The ranges of the matrices V^TC^{-1} and $C^{-1}U$ closely approximate the left and right singular spaces associated with the r smallest singular values provided the $n\times n$ nonsingular matrix A has numerical nullity r, U and V are $n\times r$ matrices, and the matrix $C=A+UV^T$ is nonsingular and well conditioned.

(b) Likewise for an $n\times n$ nonsingular matrix A having numerical rank q the ranges of the matrices $V_-^TC_-$ and C_-U_- closely approximate the left and right singular spaces of A associated with its q largest singular values provided U_- and V_- are $n\times q$ matrices, the matrix $H=I_q+V_-^TAU_-$ is nonsingular, and $C_-=A-AU_-H^{-1}V_-^TA$.

1.5 Structured input matrices and randomized augmentation

Assume a matrix A with displacement structure having a small numerical nullity r or a small numerical rank q. In this case additive preprocessing $A\Longrightarrow C=A+UV^T$ or $C_-^{-1}=A^{-1}+U_-V_-^T$ can spoil the structure a little but cannot completely destroy it: the transition $A\Longrightarrow C$ (resp. $A^{-1}\Longrightarrow C_-^{-1}$) can increase the displacement rank of A by at most $2r+1$ (resp. $2q+1$) (cf. [P01]). This enables us to extend our nearly optimal upper bounds on the number of bitwise (Boolean) operations involved into the solution of general ill conditioned linear systems of equations with matrices having small numerical nullities or ranks to the case where such matrices have displacement structure of Toeplitz or Hankel type (see [Pa]).

We can decrease the negative impact of our preprocessing on the structure of A if we endow the matrices U and V with consistent structure. E.g., we cannot extend our proof that $\kappa(C)$ has order

$\sigma_1(A)/\sigma_{n-r}(A)$ under this assumption, but such an upper bound has been in good accordance with the results of our extensive tests with highly structured and sparse matrices U and V.

We can even better preserve the input structure if we apply randomized augmentation, such as $A \Longrightarrow K = \begin{pmatrix} A & -U \\ V^T & W \end{pmatrix}$ for random structured matrices U, V and W.

Augmentation is closely linked to additive preprocessing, has similar preconditioning power, allows similar applications and enables perfect preservation of the input matrix structure (see Section 5.6).

In Sections 2.8 and 2.9) we made some initial steps towards extending formal study of conditioning of random general matrices towards random structured matrices, but we went much farther in empirical study of random Toeplitz matrices. Our tests showed that they tend to be reasonably well conditioned; this can be interesting to compare with the study of important special classes of Toeplitz matrices in [BG05].

1.6 Estimation of numerical rank and numerical nullity

One can compute numerical nullity r by means of at most $2\lceil \log_2 r \rceil$ steps of binary search whose every search step tests whether the matrix $C = A + UV^T$ is nonsingular and well conditioned for a pair of $n \times s$ random and properly scaled matrices U and V, and a candidate integer s, $s = 0, 1, 2, 4, \ldots$. Instead one can begin binary search with an upper bound $r_+ \geq r$, e.g., with $r_+ = n - 1$, and in at most $\lceil \log_2(n - r) \rceil$ steps compute r as the minimum integer for which the matrix C is nonsingular and well conditioned and the ratio $\frac{||AC^{-1}U||}{||A||\ ||C^{-1}U||}$ is small [PQ10, Algorithm 6.7].

We can begin with random scaled additive preprocessors U and V of a larger size expecting to obtain a better conditioned matrix $C = A + UV^T$, and then we can compress the preprocessors keeping smaller value of $\kappa(C)$ (see [PQ10, Flowchart 8.1]).

The power transforms $A \Longrightarrow B = (AA^H)^q A$ imply that $\sigma_j(B) = (\sigma_j(A))^{2q+1}$ and can be employed for positive integers q to increase the gaps between consecutive singular values of A.

1.7 The test results

The results of our extensive tests (the contribution of the second and the third authors) are in good accordance with our theoretical estimates. In particular we match the output accuracy of the customary algorithms but outperform them in terms of the CPU time even in the case of Toeplitz inputs (see Table 7.10). Some results of our experiments may be of independent interest, e.g., the demonstration that random Toeplitz matrices do not tend to be ill conditioned, even though users most frequently with ill conditioned Toeplitz matrices [BG05].

1.8 Related and further works

For the early works on approximation by low-rank matrices see [HMT11], [GTZ97], [GT01], [GOS08], and the bibliography therein. Preconditioning of iterative algorithms for linear systems of equations is a classical subject [A94], [B02], [G97]. The open problem of creating inexpensive preconditioners for general use has been around for a long while. Randomized preconditioning was a novel invention proposed and developed by the first author in a number of papers (see [PGMQ], [PQa], [PQZa], [PQZC], [PZ11] and the references therein). Computations with matrices having displacement structure is a classical topic, usually traced back to [KKM79]. Unification of these computations based on operating with them in terms of their displacements and the method of displacement transformation can be traced to [P90] (cf. [P01]). Treatment of ill conditioned structured matrices is a well known open challenge (cf. [VBHK01]); the best customary recipes employ displacement transformation and involve quadratic arithmetic time or large overhead constants [GKO95], [CGLX], [CGSXZ], [G98], [P10], [R06]; further advance in [Pa] relies on randomized additive preconditioning.

The contributions of the present paper include a formal proof of the power of randomized additive preconditioning for input matrices having small numerical nullities or small numerical ranks, applications to approximations with matrices of a low rank and to multiplicative preconditioning,

block factorization in Sections 5.4 and 5.5 based on approximation of trailing and leading singular spaces, and the estimates for condition numbers of circulant matrices.

As most natural directions for our further research, we consider extension of our study to augmentation and semi-augmentation in Remark 5.5, specification to structured matrices, the advance of our approximation by low-rank matrices (as we stated in Section 1.1), and its extension to tensor computations. Exploiting the link between tensor and matrix computations for their acceleration is a large fashionable subject (see, e.g., [T00], [MMD08], [OT09]). Its origin can be traced to the technique of trilinear aggregation in [P72], which was a basic ingredient of fast algorithms for matrix multiplication (see [P84], [CW90], [LPS92], and [K04]) and was the first example of the acceleration of fundamental matrix computations by means of tensor decomposition. The next highly important step, also motivated by application to fast matrix multiplication in [BCLR79], was the study of border rank for matrix and tensor decompositions [B80], [B85], [B86], [BC87].

1.9 Organization of the paper

We devote the next section to the definitions and basic results for randomized preprocessing, including the estimates from [SST06] for the condition numbers of Gaussian random matrices and our extension of these estimates to the circulant and Toeplitz cases as well as the SMW formula and our dual version of it. In Section 3 we estimate the condition numbers of randomized matrix products; the results imply support of GENP by means of randomized multiplication. In Section 4 we prove that our randomized preprocessing is expected to transform an ill conditioned matrix into a well conditioned matrix provided the input matrix has a small positive numerical nullity (with immediate extension, based on our dual SMW formula, to the case of a small positive numerical rank). In Section 5 we cover applications of the latter results to approximation by low-rank matrices, approximation of trailing and leading singular spaces, randomized additive and multiplicative preconditioning for linear systems of equations, as well as derandomization and augmentation techniques; we also comment on the extension of matrix structure in our preprocessing. In Section 6 we elaborate upon a sample application of augmentation to the solution of a Toeplitz linear system whose coefficient matrix has nullity one. In Section 7 we present the results of numerical tests, which are the contribution of the second and the third authors. In the Appendix we briefly recall Newton's iteration for the inversion of structured matrices and propose its refinement with our preconditioning techniques and its heuristic acceleration.

2 Some definitions and basic results on matrix computations

2.1 Some basic definitions

We use and extend the customary definitions in [GL96], [H02], [S98] on matrix computations and will reuse the definitions in the Introduction.

$\mathbb{R}$ and $\mathbb{C}$ are the fields of real and complex numbers, respectively.

A flop is an arithmetic operation with these numbers.

A^T and A^H denote the transpose and the Hermitian transpose of an $m \times n$ matrix A, respectively.

A^{-H} denotes the matrix $(A^{-1})^H = (A^H)^{-1}$.

A^H is A^T and A^{-H} is $A^{-T} = (A^{-1})^T = (A^T)^{-1}$ for a real matrix A.

A matrix A is Hermitian if $A = A^H$.

A matrix $A = B^H B$ is Hermitian positive definite if B is a nonsingular matrix.

$A^{(k)}$ denotes the $k \times k$ leading, that is northwestern block submatrix of a matrix A.

A matrix of a rank ρ has *generic rank profile* if all its $i \times i$ leading blocks are nonsingular for $i = 1, \dots, \rho$. If such a matrix is itself nonsingular, then it is called *strongly nonsingular*.

$(B_1 \mid \dots \mid B_k) = (B_j)_{j=1}^k$ 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 $B_1, \dots, B_k$.

I_n and I denote the $n \times n$ identity matrix $(\mathbf{e}_j)_{j=1}^n = (\mathbf{e}_1 \mid \dots \mid \mathbf{e}_n)$.

J_n and J denote the $n \times n$ reflection matrix $(\mathbf{e}_j)_{j=n}^1 = (\mathbf{e}_n \mid \dots \mid \mathbf{e}_1)$.

$O_{k,l}$ and O denote the $k \times l$ matrix filled with zeros. $\mathbf{0}_k$ is the vector $O_{k,1}$.

We drop the subscripts and write I, O, and $\mathbf{0}$ where the size of the matrix or the vector is not important or is defined by the context.

2.2 Range, null space, rank, nullity, and nmbs

$\mathcal{R}(A)$ denotes the range of the matrix A, that is the linear space $\{\mathbf{z}:\ \mathbf{z}=A\mathbf{x}\}$ generated by its columns, $\mathcal{N}(A)$ its null space $\{\mathbf{v}:\ A\mathbf{v}=\mathbf{0}\}$, $\rho=\operatorname{rank}A=\dim\mathcal{R}(A)$ its rank, and $\operatorname{nul}A=\dim\mathcal{N}(A)$ its nullity. $\mathbf{v}$ is its null vector if $A\mathbf{v}=\mathbf{0}$.

Suppose a matrix B has full column rank and $\mathcal{R}(B)=\mathcal{N}(A)$. Then we call B a *null matrix basis* or a *nmb* for a matrix A and write $B=\operatorname{nmb}(A)$.

The nullity, the null space, null vectors, and nmbs of the transposed matrix A^T are said to be the left nullity, the left null space, left null vectors, and left nmbs of a matrix A, respectively.

2.3 Orthogonalization, norms, and SVD

A matrix $X=A^{(I)}$ is a left or right inverse of a matrix A if $XA=I$ or $AX=I$, respectively, so that $A^{(I)}=A^{-1}$ for a nonsingular matrix A.

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

$$\max_{i,j=1}^{m,n}|a_{i,j}|\le||A||=||A^H||\le\sqrt{mn}\ \max_{i,j=1}^{m,n}|a_{i,j}|, \tag{2.1}$$

$$||A||^2\le||A||_1||A||_\infty. \tag{2.2}$$

$\mathbf{v}$ is a *unit* or *normalized* vector if $||\mathbf{v}||=\sqrt{\mathbf{v}^H\mathbf{v}}=1$.

An $m\times n$ matrix U (for $m\le n$) is *unitary* or *orthonormal* if $U^HU=I$.

QR factorization $A=QR$ of a matrix A into the product of a unitary matrix Q and an upper triangular matrix R is unique if the factor R is a square matrix with positive diagonal entries [GL96, Theorem 5.2.2]. In this case we write $Q=Q(A)$ and $R=R(A)$.

$A=S_A\Sigma_AT_A^H$ is an *SVD* or *full SVD* of an $m\times n$ matrix A of a rank ρ provided $S_AS_A^H=S_A^HS_A=I_m$, $T_AT_A^H=T_A^HT_A=I_n$, $\Sigma_A=\operatorname{diag}(\widehat{\Sigma}_A,O_{m-\rho,n-\rho})$, $\widehat{\Sigma}_A=\operatorname{diag}(\sigma_j(A))_{j=1}^{\rho}$, $\sigma_j=\sigma_j(A)=\sigma_j(A^H)$ is the jth largest singular value of a matrix A. These values have the minimax characterization

$$\sigma_j=\max_{\dim(\mathbb{S})=j}\ \min_{\mathbf{x}\in\mathbb{S},\ ||\mathbf{x}||=1}\ ||A\mathbf{x}||,\ j=1,\dots,\rho, \tag{2.3}$$

where $\mathbb{S}$ denotes linear spaces [GL96, Theorem 8.6.1]. They turn into zero, $\sigma_j=0$, where $j>\rho$. It follows that $\sigma_j(A)$ is the distance from the matrix A to the nearest matrix of a rank $j-1$ for $j=1,\dots,\rho+1$,

$$\sigma_1=\max_{||\mathbf{x}||=1}||A\mathbf{x}||=||A||\ \text{and if}\ m\ge n,\ \text{then}\ \sigma_n=\min_{||\mathbf{x}||=1}||A\mathbf{x}||. \tag{2.4}$$

The minimax characterization (2.3) implies

Fact 2.1. *Fix two positive integers p and q and assume that A_0 is a $p\times q$ submatrix of a matrix A. Then $\sigma_j(A)\ge\sigma_j(A_0)$ for $j=1,2,\dots,\min\{p,q\}$.*

If $\sigma_q>\sigma_{q+1}$, in which case $q\le\rho$, then the first q columns of the matrices S_A and T_A generate the leading left and right singular spaces $\mathbb{S}_A^{(q)}=\mathcal{R}(S_A(I_q\mid O_{q,m-q})^T)$ and $\mathbb{T}_A^{(q)}=\mathcal{R}(T_A(I_q\mid O_{q,n-q})^T)$, respectively, associated with the q largest singular values of the matrix A. The orthogonal complements $\mathbb{S}_{A,m-q}$ and $\mathbb{T}_{A,n-q}$ of these singular spaces are the left and right trailing singular spaces associated with the $m-q$ and $n-q$ smallest singular values of the matrix A, respectively.

$\Sigma_A^+=\operatorname{diag}((\widehat{\Sigma}_A)^{-1},O_{n-\rho,m-\rho})$ and $A^+=T_A\Sigma_A^+S_A^H$ are the Moore–Penrose generalized (or pseudo) inverses of the matrices Σ_A and A, respectively. A^+ is a left or right inverse of a matrix A of full rank. $A^+=A^{-1}$ for a nonsingular matrix A. $||A^+||=1/\sigma_\rho(A)$ for a matrix A of rank ρ.

2.4 Condition number, perturbation norm bounds, and numerical rank and nullity

Hereafter the concepts "large", "small", "near", "closely approximate", "ill conditioned" and "well conditioned" are quantified in the context.

For two positive parameters a and b we write $a \gg b$ and $b \ll a$ if the ratio a/b is large and write $a \approx b$ if the ratio is close to one or if $b = 0$ and $|a|$ is small. For two matrices A and B we write $A \approx B$ if $||A - B|| \ll ||A||$.

An $m \times n$ matrix A has *numerical rank* ρ if the ratios $\sigma_j(A)/||A||$ are small for $j > \rho$ but are not small for $j \leq \rho$. In this case the matrix has *numerical nullity* $n - \rho$.

Remark 2.1. *Unlike the nullity and the rank, numerical nullity and numerical rank are not well defined for a large class of ill conditioned matrices, in particular for all matrices A having nested clusters of small singular values but also for the matrix class represented by a 1000×1000 matrix A with singular values $\sigma_j(A) = 2^{1000-j}$, $j = 1, 2, \dots, 1000$, e.g., by* $\operatorname{diag}(2^{1000-j})_{j=1}^{1000}$.

$\kappa(A) = \frac{\sigma_1(A)}{\sigma_\rho(A)} = ||A||\ ||A^+||$ is the condition number of a matrix A of a rank ρ. Such a 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], and [S98, Section 5.3] on effective estimation of norms and condition numbers. $\kappa(A) = ||A||\ ||A^{-1}||$ for a nonsingular matrix A.

A map of matrices $A \Longrightarrow B$ is called *preconditioning* if $\kappa(B) \ll \kappa(A)$ and if the solution of a linear system $A\mathbf{y} = \mathbf{b}$ is readily reduced to the solution of a linear system $B\mathbf{x} = \mathbf{f}$, e.g., if $B = AM$ for a readily computable matrix M. Preconditioning of the coefficient matrix accelerates convergence of CG, GMRES and other most popular iterative algorithms to the solution of a linear system $B\mathbf{x} = \mathbf{f}$, but the convergence is particularly fast where the matrix $I - B$ has a small numerical rank, even if the condition number $\kappa(B)$ is large [A94], [B02], [G97]. This motivates the following definition: M is a *multiplicative preprocessor of a level* r for a square matrix A if the matrix $MA - I$ or $AM - I$ has a numerical rank at most r. Clearly $M = A^{-1}$ is a multiplicative preprocessor of level zero, but it is assumed that the computation of a preprocessor is substantially simpler than inversion.

The minimax characterization (2.3) implies that an $m \times n$ matrix A of full rank is ill conditioned if and only if it is close to a rank deficient matrix B, such that the ratio $||A - B||/||A||$ is small.

By extending the definitions in Section 2.1 we say that a matrix having a numerical rank ρ has *generic conditioning profile* if all its $i \times i$ leading blocks are well conditioned for $i = 1, \dots, \rho$. If such a matrix is itself well conditioned, then it is *strongly well conditioned*.

One can readily verify the following property (see [PQZa]).

Theorem 2.1. *Gaussian elimination with no pivoting applied to a matrix A involves no divisions by zeros (resp. by values that are absolutely small relative to the norm $||A||$) while computing triangular factorizations of the leading block submatrices $A^{(j)}$ for $j = 1, \dots, \rho$, for ρ denoting* $\operatorname{rank} A$ *(resp. numerical rank of A) if and only if the matrix A has generic rank (resp. generic conditioning) profile.*

Similar property holds for block Gaussian elimination (see [PQZa]) and can be extended to the MBA algorithm of [M80] and [BA80].

These results motivate the task of *generic preconditioning* whose goal is the transformation of a well conditioned matrix into a matrix having generic conditioning profile (see Corollary 3.4).

The next theorem bounds the perturbation norm of the solution of a linear system of equations in terms of its input perturbation norms and the condition number $\kappa(A)$. As the bound ϵ on the input perturbation norm converges to zero, the bound on the output perturbation norm has order $2\epsilon\kappa(A)$.

Theorem 2.2. *[H02, Section 7.1, page 121]. Let $A\mathbf{y} = \mathbf{b}$ and $(A + \Delta(A))\tilde{\mathbf{y}} = \mathbf{b} + \Delta(\mathbf{b})$ for a pair of nonsingular matrices A and $A + \Delta(A)$ and two vectors $\mathbf{b}$ and $\Delta(\mathbf{b})$ such that $||\Delta(A)|| \leq \epsilon||A||$, $||\Delta(\mathbf{b})|| \leq \epsilon||\mathbf{b}||$ for $\epsilon\kappa(A) < 1$. Then $\frac{||(\tilde{\mathbf{y}} - \mathbf{y})||}{||\tilde{\mathbf{y}}||} \leq 2\epsilon\frac{\kappa(A)}{1 - \epsilon\kappa(A)}$.*

Backward analysis of rounding errors extends this result as follows [GL96, Section 3.4.6], [H02, Theorems 19.5], [S98, Theorem 3.4.9].

Corollary 2.1. *Gaussian elimination with pivoting uses $\frac{2}{3}n^3 + O(n^2)$ flops with rounding to a precision p to produce an approximate solution $\tilde{\mathbf{y}}$ to a nonsingular linear system $A\mathbf{y} = \mathbf{b}$ of n equations and an approximate inverse $X \approx A^{-1}$ such that $\frac{||\tilde{\mathbf{y}}-\mathbf{y}||}{||\mathbf{y}||} = O(2^{-p}n^2\kappa(A \mid \mathbf{b}))$ and $\frac{||X-A^{-1}||}{||A^{-1}||} = O(2^{-p}n^2\kappa(A))$ as $p \to \infty$.*

Now suppose an $n \times n$ matrix $\tilde{A}$ has a rank ρ and nullity $r = n - \rho$. If such a matrix is well conditioned, then all its sufficiently close neighbours A have numerical rank ρ and numerical nullity r, that is have exactly r singular values that are small relative to the norm $||\tilde{A}||$, even though almost all of these neiboughrs have rank n and nullity zero. The minimax characterization implies that conversely, every $n \times n$ matrix A having a positive numerical nullity r is close to a well conditioned singular matrix $\tilde{A}$ of rank $n-r$. The range $\mathcal{R}(\tilde{A})$ of the latter matrix approximates its *leading singular space* $\mathbb{T}_A^{(n-r)}$ associated with the $n-r$ largest singular values; the null space $\mathcal{N}(\tilde{A}))$ approximates the *trailing singular space* $\mathbb{T}_{A,r}$ associated with the r smallest singular values.

r is also the left numerical nullity of such a matrix A, whose left leading and trailing singular spaces $\mathbb{S}_A^{(n-r)}$ and $\mathbb{S}_{A,r}$ associated with the $n-r$ largest and r smallest singular values, respectively, are approximated by the range and the null space of the matrix $\tilde{A}^H$, respectively.

2.5 The SMW and dual SMW formulae

Theorem 2.3. *Suppose $A \in \mathbb{C}^{n\times n}$, $U, V \in \mathbb{C}^{n\times r}$, the matrix $C = A + UV^H$ is nonsingular, $G = I_r - V^H C^{-1}U$, and $0 < r < n$. (The matrix G is called the Gauss transform and the Schur complement [GL96]). Then we have factorizations*

$$\begin{pmatrix} C & U \\ V^H & I_r \end{pmatrix} = \begin{pmatrix} I_n & U \\ O_{r,n} & I_r \end{pmatrix} \begin{pmatrix} A & O_{n,r} \\ O_{r,n} & I_r \end{pmatrix} \begin{pmatrix} I_n & O_{n,r} \\ V^H & I_r \end{pmatrix} \tag{2.5}$$

and

$$\begin{pmatrix} C & U \\ V^H & I_r \end{pmatrix} = \begin{pmatrix} I_n & O_{n,r} \\ V^H C^{-1} & I_r \end{pmatrix} \begin{pmatrix} C & O_{n,r} \\ O_{r,n} & G \end{pmatrix} \begin{pmatrix} I_n & C^{-1}U \\ O_{r,n} & I_r \end{pmatrix}. \tag{2.6}$$

Furthermore if the matrix A is nonsingular, then so is the matrix G, and we have the Sherman–-Morrison–Woodbury formula (hereafter referred to as the SMW formula) $A^{-1} = (C - UV^H)^{-1} = C^{-1} + C^{-1}UG^{-1}V^HC^{-1}$.

Proof. The claimed factorizations and nonsigularity are readily verified. The SMW formula of [GL96, page 50], [S98, Corollary 4.3.2] follows if we invert factorization (2.5). □

Corollary 2.2. *Suppose $A \in \mathbb{C}^{n\times n}$, $U, V \in \mathbb{C}^{n\times r}$, and the matrices A and $C = A + UV^H$ are nonsingular. Then the matrix C^{-1} is a multiplicative preprocessor of level r for the matrix A.*

We have the straightforward bound

$$||G|| \le ||V||\ ||C^{-1}||\ ||U|| + 1. \tag{2.7}$$

Furthermore by inverting factorization (2.6) we obtain that

$$||G^{-1}|| \le \max\{||A^{-1}||, 1\}(||V||\ ||C^{-1}|| + 1)(||U||\ ||C^{-1}|| + 1)(||U|| + 1)(||V|| + 1). \tag{2.8}$$

Next apply the SMW formula to the matrix A^{-1} and a pair of matrices $U_-, V_- \in \mathbb{C}^{n\times q}$ for $0 < q < n$ and obtain the *dual SMW formula*

$$A^{-1} = (C_-)^{-1} - U_-V_-^H \text{ for } C_- = A - AU_-H^{-1}V_-^HA \text{ and } H = I_q + V_-^HAU_- \tag{2.9}$$

provided that H and C_- are nonsingular matrices.

Both SMW and dual SMW formulae can be extended to the case of rectangular matrices A [PQ10]. In the case of square matrices A they imply that

$$\det A = (\det C)\det G = (\det C_-)/\det H. \tag{2.10}$$

To deduce the former equation, equate the right hand sides of equations (2.5) and (2.6) and recall that the determinant of a unit triangular matrix equals one. To deduce the latter equation in (2.10) first extend the former equation to the expression $A^{-1} = (C_-)^{-1} - U_- V_-^H$ in (2.9) replacing $C = A + UV^H$. This yields the expression $\det(A^{-1}) = (\det(C_-)^{-1}) \det H$. It remains to substitute the equation $\det(M^{-1}) = (\det M)^{-1}$ for $M = A$ and $M = C_-$.

2.6 Random sampling and random matrices

$|\Delta|$ is the cardinality of a set Δ in any fixed ring. *Random sampling* of elements from a set Δ is their selection from this set at random and independently of each other. A matrix is *random* if its entries are randomly sampled from a fixed set Δ. Random sampling is *uniform* if it is done under the uniform probability distribution on the set Δ.

Recall that 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 2.1. *[DL78], [S80], [Z79]. For a set Δ of 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.*

Lemma 2.1 implies that a fixed nonvanishing polynomial vanishes with a probablity converging to zero if the values of its variables are sampled under any reasonable probability distribution on the set Δ whose cardinality converges to infinity. Under the uniform probability distribution the probability is estimated most readily.

Corollary 2.3. *Under the assumptions of Lemma 2.1 let the values of the variables of the polynomial be randomly and uniformly sampled from the set Δ. Then the polynomial vanishes with a probability at most $\frac{d}{|\Delta|}$.*

Corollary 2.4. *Let the entries of an $m \times n$ 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 singular with a probability at least $1 - \frac{k}{|\Delta|}$ and (b) is strongly nonsingular with a probability at least $1 - \sum_{i=1}^{k} \frac{i}{|\Delta|} = 1 - \frac{(k+1)k}{2|\Delta|}$. Furthermore (c) if the submatrix M is indeed nonsingular, then any entry of its inverse is nonzero with a probability at least $1 - \frac{k-1}{|\Delta|}$.*

Proof. The claimed bounds 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. This 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 a $(k-1) \times (k-1)$ subdeterminant of the input M, and so it is a polynomial of degree $k-1$ in its entries. □

Definition 2.1. *$F_X(y) = Probability\{X \le y\}$ for a real random variable X is the* cumulative distribution function (CDF) *of X evaluated at y. $F_A(y) = F_{\sigma_l(A)}(y)$ for an $m \times n$ matrix A and an integer $l = \min\{m, n\}$. A matrix is a* Gaussian random matrix *with a mean μ and a variance σ^2 if it is filled with independent Gaussian random variables, all having the same mean μ and variance σ^2. $\mathcal{G}_{\mu,\sigma}^{m \times n}$ denotes the set of such $m \times n$ Gaussian random matrices. For $\mu = 0$ and $\sigma^2 = 1$ they turn into* standard Gaussian random matrices. *$F_{\mu,\sigma}(y) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{y} \exp(-\frac{(x-\mu)^2}{2\sigma^2})dx$ for a Gaussian random variable with a mean μ and a variance σ^2. $\chi_{\mu,\sigma,n}(y)$ is the CDF of the random function $(\sum_{i=1}^{n} Y_i^2)^{1/2} = ||(Y_i)_{i=0}^n||$ where $Y_1, \dots, Y_n$ are n independent Gaussian random variables sharing a mean μ and a variance σ^2, that is where $(Y_i)_{i=0}^n$ is Gaussian random vector with a mean μ and a variance σ^2. For $y \ge 0$ we have $\chi_{0,1,n}(y) = \frac{2}{2^{n/2}\Gamma(n/2)} \int_{-\infty}^{y} x^{n-1} \exp(-x^2/2)dx$ where $\Gamma(h) = \int_0^\infty x^{h-1} \exp(-x)dx$, $\Gamma(n+1) = n!$ for nonnegative integers n.*

2.7 Conditioning of Gaussian random matrices

Gaussian random matrices in Definition 2.1 tend to be well conditioned [D88], [E88], [CD05]. We will study them based on the estimates of [SST06], [ST02]; we recall them below. The estimates in [E88] and [CD05] have smaller overhead constants and in the case of rectangular nonsquare matrices are superior by order of magnitude, but [SST06], [ST02] prove that even the sum of two matrices $M \in \mathbb{R}^{m\times n}$ and $W \in \mathcal{G}_{\mu,\sigma}^{m\times n}$ is expected to be well conditioned as long as the ratio $\sigma/||M||$ is not small.

The following upper bound on the probability that for a Gaussian random matrix W the smallest singular value of a matrix $A = W + M$ is less than a scalar y can also be viewed as a probabilistic lower bound on the smallest singular value of the matrix A.

Theorem 2.4. *Suppose $M \in \mathbb{R}^{m\times n}$ is a fixed matrix, $W \in \mathcal{G}_{\mu,\sigma}^{m\times n}$, $A = W + M$, $l = \min\{m, n\}$, and $y \geq 0$. Then $F_A(y) \leq 2.35\, y\sqrt{l}/\sigma$.*

Proof. Clearly the matrix $A = M + W$ has full rank with probability one because W is a Gaussian random matrix. In view of Fact 2.1 it is sufficient to prove the theorem in the case where $m = n$, and in this case the theorem turns into [SST06, Theorem 3.3]. □

The two following theorems supply lower bounds on the probabilities that $||W|| \leq y$ and $\kappa(A) \leq y$ for a scalar y, $A = W + M$, and a Gaussian random matrix W. They can also be viewed as probabilistic upper bounds on the norm $||W||$ and the condition number $\kappa(A)$.

Theorem 2.5. *(See [DS01, Theorem II.7].) Suppose $W \in \mathcal{G}_{0,\sigma}^{m\times n}$, $l = \min\{m, n\}$ and $y \geq 2\sigma\sqrt{l}$. Then $F_{||W||}(y) \geq 1 - \exp(-(y - 2\sigma\sqrt{l})^2/(2\sigma^2))$.*

Theorem 2.6. *(See [SST06, Theorem 3.1].) Under the assumptions of Theorem 2.4, let $||M|| \leq \sqrt{l}$, $\mu = 0$, $\sigma \leq 1$. Then $F_{\kappa(A)}(y) \geq 1 - (14.1 + 4.7\sqrt{(2\ln y)/n})n/(y\sigma)$ for all $y \geq 1$.*

On a further improvement of this bound by a factor $\sqrt{\log n}$, see [W04].

Theorem 2.6 is deduced in [SST06] based on combining Theorems 2.4 and 2.5. The proof of the former theorem relies on the two lemmas below that we use in the next subsections.

Lemma 2.2. *Suppose y is a positive number, $\mathbf{w} \in \mathbb{R}^{n\times 1}$ is any fixed real unit vector, $||\mathbf{w}|| = 1$, $M \in \mathbb{R}^{n\times n}$ is a fixed real matrix, $W \in \mathcal{G}_{\mu,\sigma}^{n\times n}$, and $A = W + M$. Let Q be a unitary matrix such that $Q\mathbf{w} = \mathbf{e}_1$ and let $B = QA = (\mathbf{b}_1 \mid \ldots \mid \mathbf{b}_n)$. Then*

$$\text{Probability}\{||A^{-1}\mathbf{w}|| > y\} \leq \max_{\mathbf{b}_1,\ldots,\mathbf{b}_n} \text{Probability}\{|\mathbf{t}^T\mathbf{b}_1| < 1/y\}$$

where $||\mathbf{t}|| = 1$ and $\mathbf{t}^T\mathbf{b}_i = 0$ for $i = 2, \ldots, n$.

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

Lemma 2.3. *[SST06, Lemma A.2]. For a positive δ, a vector $\mathbf{b} \in \mathcal{G}_{\mu,\sigma}^{n\times 1}$ and a unit real vector $\mathbf{t}$, we have* $\text{Probability}\{|\mathbf{t}^T\mathbf{b}| < \delta\} \leq \sqrt{\frac{2}{\pi}}\frac{\delta}{\sigma}$.

Remark 2.2. *The latter bound is independent of n, and so it holds even if all coordinates of the vector $\mathbf{b}$ are fixed, except for one coordinate in $\mathcal{G}_{\mu,\sigma}$.*

2.8 Conditioning of Gaussian random Toeplitz matrices

$m \times n$ Toeplitz matrix $T = (t_{i-j})_{i,j=1}^{m,n}$ is defined by the vector $(t_h)_{h=1-n,2-n,\ldots,m-1}$ of dimension $m + n - 1$, made up of the entries of the first row and column of T. If this is a Gaussian random vector with a mean μ and a variance σ^2, then we write $T \in \mathcal{T}_{\mu,\sigma}^{m\times n}$ and call T a *Gaussian random Toeplitz matrix* with a mean μ and a variance σ^2.

An $n \times n$ lower triangular Toeplitz matrix $T = (t_{i-j})_{i,j=1}^{n}$ where $t_k = 0$ for $k < 0$ is completely defined by the vector its first column $T\mathbf{e}_1$; hereafter $Z(\mathbf{v})$ denotes such a matrix with the first column

$\mathbf{v} = Z(\mathbf{v})\mathbf{e}_1$. $Z = Z(\mathbf{e}_2)$ is the $n \times n$ downshift matrix whose all entries are zeros except for the first subdiagonal filled with ones; $Z\mathbf{v} = (v_i)_{i=0}^{n-1}$ for a vector $\mathbf{v} = (v_i)_{i=1}^{n}$ and $v_0 = 0$; furthermore $Z^n = O$, $Z(\mathbf{v}) = \sum_{i=0}^{n-1} v_{i-1}Z^i$.

Observe that $||Z(\mathbf{v})||_1 = ||Z(\mathbf{v})||_\infty = ||\mathbf{v}||_1$. By combining these equations with bound (2.2) obtain that

$$||Z(\mathbf{v})|| \le ||\mathbf{v}||_1. \tag{2.11}$$

The celebrated formula of [GS72] expresses the inverse $X = (x_{ij})_{i,j=1}^{n}$ of a nonsingular $n \times n$ Toeplitz matrix through its first column and $\mathbf{x}_1 = X\mathbf{e}_1$ its last column $\mathbf{x}_n = X\mathbf{e}_n$ as follows,

$$x_{11}X = Z(\mathbf{x}_1)Z^T(J\mathbf{x}_n) - Z(Z\mathbf{x}_n)Z^T(ZJ\mathbf{x}_1) \tag{2.12}$$

provided $x_{11} \neq 0$.

It is easy to deduce that with probability one a Toeplitz matrix $T \in \mathcal{T}_{\mu,\sigma}^{n\times n}$ is nonsingular and the entry $x_{11} = \mathbf{e}_1T^{-1}\mathbf{e}_1 = \det T_{11}/\det T$ of its inverse does not vanish. Here T_{11} is the $(n-1)\times(n-1)$ submatrix of T obtained by deleting the first row and the first column.

Furthermore the proof of [SST06, Lemma 3.2] can be readily extended to deduce

Lemma 2.4. *Suppose y is a positive number, $T \in \mathcal{T}_{\mu,\sigma}^{n\times n}$, j is an integer, $1 \le j \le n$, $\bar{\mathbf{x}}_j \in \mathbb{R}^{n\times 1}$ is the unit vector orthogonal to all vectors $T\mathbf{e}_i$ for $i \neq j$, $M \in \mathbb{R}^{n\times n}$ is a fixed real matrix, and $A = W + M$. Then*

$$\text{Probability}\{||T^{-1}\mathbf{e}_j|| > y\} \le \text{Probability}\{|\bar{\mathbf{x}}_j^T T\mathbf{e}_j| < 1/y\}.$$

Now observe that each of the column vectors $T\mathbf{e}_1$ and $T\mathbf{e}_n$ has an entry not shared with the other entries of T, recall Remark 2.2 and deduce that

$$\text{Probability}\{|\bar{\mathbf{x}}_j^T T\mathbf{e}_j| < \delta\} \le \sqrt{\frac{2}{\pi}}\frac{\delta}{\sigma}$$

for $j = 1$ and $j = n$.

By combining these estimates with equation (2.12) and bound (2.11) deduce the following result.

Theorem 2.7. *Assume that $T \in \mathcal{T}_{\mu,\sigma}^{n\times n}$ and $y \ge 0$. Then with probability one the matrix T is nonsingular and the entry $x_{11} = \mathbf{e}_1T^{-1}\mathbf{e}_1 = \det T_{11}/\det T$ of its inverse $X = (x_{ij})_{i,j=1}^{n}$ does not vanish; furthermore $||x_{11}X|| \le 2RS$ for two random variables R and S such that $F_R(y) \le 2.35\sqrt{n}/(\sigma y)$ and $F_S(y) \le 2.35\ y\sqrt{n}/(\sigma y)$.*

The theorem provides probabilistic upper bound on the product $||x_{11}X||$ but the factor $|x_{11}|$ is not so easily estimated from below. Our extensive tests, however, show that random Toeplitz matrices tend to be reasonably well conditioned, unlike Toeplitz matrices of some important special classes (cf. [BG05]).

2.9 Conditioning of Gaussian random circulant matrices

$Z_f = Z + f\mathbf{e}_n^T\mathbf{e}_1$ is the unit f-circulant matrix. An *f-circulant matrix* $Z_f(\mathbf{v}) = \sum_{i=0}^{n-1} v_i Z_f^i = (z_{i-j \mod n})_{i,j=1}^{n}$ is an $n \times n$ Toeplitz matrix defined by its first column vector $\mathbf{v} = (v_i)_{i=0}^{n-1}$ and a scalar $f \neq 0$. We call such a matrix *Gaussian random f-circulant matrix* with a mean μ and a variance σ^2 if $\mathbf{v} \in \mathcal{G}_{\mu,\sigma}^{n\times 1}$.

f-circulant matrix is called *circulant* if $f = 1$ and *skew circulant* if $f = -1$. By replacing f by zero we can arrive at the lower triangular Toeplitz matrices $Z(\mathbf{v})$.

Theorem 2.8. *(See [CPW74].) We have $Z_1(\mathbf{v}) = \Omega^{-1}D(\Omega\mathbf{v})\Omega$. More generally, for any $f \neq 0$, we have $Z_{f^n}(\mathbf{v}) = U_f^{-1}D(U_f\mathbf{v})U_f$ where $U_f = \Omega D(\mathbf{f})$, $\mathbf{f} = (f^i)_{i=0}^{n-1}$, $D(\mathbf{u}) = \text{diag}(u_i)_{i=0}^{n-1}$ for a vector $\mathbf{u} = (u_i)_{i=0}^{n-1}$, and $\Omega = (\omega_n^{ij})_{i,j=0}^{n-1}$ is the $n \times n$ matrix of the discrete Fourier transform at n points, $\omega_n = \exp(2\pi\sqrt{-1}/n)$ being a primitive n-th root of one.*

The theorem implies that multiplication and inversion of f-circulant matrices (wherever feasible) produce f-circulant matrices and can be performed in $O(n \log n)$ flops for $n \times n$ inputs based on FFT.

Next we estimate the norms of a random Gaussian f-circulant matrix and its inverse.

Theorem 2.9. *Assume an $n \times n$ circulant matrix $A = Z_1(\mathbf{v}) = \sum_{i=0}^{n-1} v_i Z_1^i$ for $\mathbf{v} \in \mathcal{G}_{\mu,\sigma}^{n\times 1}$. Then we have a) $F_{||A||}(y) \geq \chi_{\mu,\sigma,n}(y/\sqrt{n})$ and b) $F_{\sigma_n(A)}(y) \leq \sqrt{\frac{2}{\pi}}\frac{1}{\sigma y n}$ for all nonnegative y and for $\chi_{\mu,\sigma}(y)$ in Definition 2.1.*

Proof. Represent the matrix A as in Theorem 2.8 and write $B = \Omega A \Omega^{-1} = D(\Omega \mathbf{v})$, $\mathbf{u} = (u_i)_{i=0}^{n-1} = \Omega \mathbf{v}$. $\frac{1}{\sqrt{n}}\Omega$ is a unitary matrix, and so $\sigma_j(A) = \sigma_j(B)$ for all j, $||\mathbf{u}|| = \sqrt{n}||\mathbf{v}||$. Therefore $||A|| = ||B|| = \max_j |u_j| \leq ||\mathbf{u}|| = \sqrt{n}||\mathbf{v}||$, and we obtain part a) of the theorem, because $B = \mathrm{diag}(u_i)_{i=0}^{n-1}$ and $\mathbf{v} \in \mathcal{G}_{\mu,\sigma}^{n\times 1}$.

Furthermore we combine the equations $u_i = \mathbf{e}_i^T \Omega \mathbf{v}$ and the bounds $||\Re(\mathbf{e}_i^T \Omega)|| \geq 1$ for all i, with Lemma 2.3 and deduce that $F_{|\Re(u_i)|}(y) \leq \sqrt{\frac{2}{\pi}}\frac{1}{\sigma y}$ for any i, $i = 1, \ldots, n$. We have $F_{\sigma_n(B)}(y) = F_{\min_i |u_i|}(y)$ because $B = \mathrm{diag}(u_i)_{i=0}^{n-1}$. Clearly $|u_i| \geq |\Re(u_i)|$, and part b) of the theorem follows. □

Remark 2.3. *Our extensive experiments (cf. Table 7.3) suggest that both lower and upper estimates of Theorem 2.9 are overly pessimistic.*

Combining Theorem 2.8 with minimax characterization implies that

$$\frac{1}{g(f)}\sigma_j(Z_1(\mathbf{v})) \leq \sigma_j(Z_f(\mathbf{v})) \leq g(f)\sigma_j(Z_1(\mathbf{v}))$$

for all vectors $\mathbf{v}$, scalars $f \neq 0$, $g(f) = \max\{1, |f|\}\ \max\{1, \frac{1}{|f|}\}$, and $j = 1, 2, \ldots, n$. This enables us to extend the estimates of Theorem 2.9 to f-circulant matrices for $f \neq 0$. In particular these estimates do not change in the case of skew circulant matrices (for which $f = -1$) and show that $n \times n$ f-circulant matrices tend to be well conditioned for any fixed $f \neq 0$.

2.10 Extension to the case of complex inputs

For simplicity we assume real random matrices, but next show some basic results for the extension of our study to the case of complex matrices. According to the analysis and experiments in [E88] their tendency to be well conditioned is stronger than in the case of real matrices.

Definition 2.2. *A complex Gaussian random value with a mean μ and a variance σ is the sum $a + b\sqrt{-1}$ where a and b are a pair of independent real Gaussian random values, each having a mean μ and a variance σ. An $m \times n$ complex Gaussian random matrix with a mean μ and a variance σ is the matrix $A + B\sqrt{-1}$ where A and B are a pair of Gaussian random matrices in $\mathcal{G}_{\mu,\sigma}^{m\times n}$ having $2mn$ independent entries overall. Such matrices form the set $\mathcal{G}_{\mathbb{C},\mu,\sigma}^{m\times n}$.*

We can immediately extend Theorem 2.5 to the matrices in $\mathcal{G}_{\mathbb{C},\mu,\sigma}^{m\times n}$ because $||A + B\sqrt{-1}|| \leq ||A|| + ||B||$.

Next we extend Lemma 2.3.

Lemma 2.5. *The bound of Lemma 2.3 also holds provided $\mathbf{t} = \mathbf{r} + \mathbf{q}\sqrt{-1}$ is a fixed complex unit vector and $\mathbf{b} = \mathbf{f} + \mathbf{g}\sqrt{-1} \in \mathcal{G}_{\mathbb{C},\mu,\sigma}^{n\times 1}$ is a complex vector such that $\mathbf{f}$, $\mathbf{g}$, $\mathbf{r}$ and $\mathbf{q}$ are real vectors, $||\mathbf{t}|| = 1$, and vectors $\mathbf{f}$ and $\mathbf{g}$ are in $\mathcal{G}_{\mu,\sigma}^{n\times 1}$.*

Proof. Note that $\mathbf{t}^H\mathbf{b} = \mathbf{r}^T\mathbf{f} - \mathbf{q}^T\mathbf{g} + (\mathbf{r}^T\mathbf{g} + \mathbf{q}^T\mathbf{f})\sqrt{-1}$, and so $|\mathbf{t}^H\mathbf{b}|^2 = |\mathbf{r}^T\mathbf{f} - \mathbf{q}^T\mathbf{g}|^2 + |\mathbf{r}^T\mathbf{g} + \mathbf{q}^T\mathbf{f}|^2$. Hence $|\mathbf{t}^H\mathbf{b}| \geq |\mathbf{r}^T\mathbf{g} + \mathbf{q}^T\mathbf{f}| = |\mathbf{u}^T\mathbf{v}|$ where $\mathbf{u}^T = (\mathbf{r}^T \mid \mathbf{q}^T)$ and $\mathbf{v}^T = (\mathbf{g}^T \mid \mathbf{f}^T)$, so that $\mathbf{v} \in \mathcal{G}_{\mu,\sigma}^{1\times 2n}$ and $||\mathbf{u}|| = ||\mathbf{t}|| = 1$. It remains to apply the bound of Lemma 2.3 to real vectors $\mathbf{u}$ and $\mathbf{v}$ replacing $\mathbf{b}$ and $\mathbf{t}$, respectively. □

By combining Lemmas 2.2 and 2.5 we obtain the following extension of [SST06, Lemma 3.2] to the case of complex inputs.

Corollary 2.5. *Suppose y is a positive number, $\mathbf{w} \in \mathbb{C}^{n\times 1}$ is any fixed complex unit vector, $||\mathbf{w}|| = 1$, $M \in \mathbb{C}^{n\times n}$ is a fixed matrix, $W \in \mathcal{G}^{n\times n}_{\mathbb{C},\mu,\sigma}$, and $A = W + M$. Then*

$$\text{Probability}\{||A^{-1}\mathbf{w}|| > y\} \le \frac{2}{y\sigma\sqrt{\pi}}\ .$$

Proof. Lemmas 2.2 and 2.5 together imply the corollary for any fixed unit real vector $\mathbf{w} \in \mathbb{R}^{n\times 1}$ and the upper bound decreased by a factor $\sqrt{2}$. We extend the latter bound to the one of Corollary 2.5 follows because $\max\{||\mathbf{u}||, ||\mathbf{v}||\} \ge 1/\sqrt{2}$ provided $\mathbf{w} = \mathbf{u} + \mathbf{v}\sqrt{-1}$, $\mathbf{u} = \Re(\mathbf{w})$, $\mathbf{v} = \Im(\mathbf{w})$, and $||\mathbf{w}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2 = 1$. □

Corollary 2.6. *Suppose y is a positive number, $M \in \mathbb{C}^{n\times n}$ is a fixed matrix, $W \in \mathcal{G}^{n\times n}_{\mathbb{C},\mu,\sigma}$, and $A = W + M$. Then*

$$\text{Probability}\{||A^{-1}|| > y\} \le \frac{2n}{y\sigma\sqrt{\pi}}.$$

Proof. Recall that $||B|| = \max_{j=1}^{n} ||B\mathbf{e}_j||$ for any $n \times n$ matrix B, in particular for $B = A^{-1}$. It remains to apply Corollary 2.5 to the vectors $\mathbf{w} = \mathbf{e}_j$ for $j = 1, \dots, n$ and deduce that

$$\text{Probability}\{\max_{j=1,\dots,n} ||A^{-1}\mathbf{e}_j|| > y\} \le \frac{2n}{y\sigma\sqrt{\pi}}.$$

□

Based on the latter result, one can readily extend Theorems 2.4, 2.7 and 2.9 to the case of complex inputs.

Remark 2.4. *Corollary 2.6 extends [SST06, Theorem 3.3] to the case of complex matrices but increases the bound of the theorem by a factor $\sqrt{2n}$. We recall that random complex matrices tend to be a little better conditioned than random real matrices according to the estimates and tests in [E88], and so we conjecture that the estimated increase by a factor $\sqrt{2n}$ is overly pessimistic.*

3 Conditioning of randomized matrix products and generic preconditioning

Next we extend the estimates of Theorem 2.4 to yield probabilistic lower bounds on the smallest singular values of the products of fixed and random matrices. We begin with four lemmas. The first two of them easily follow from minimax characterization (2.3).

Lemma 3.1. *Suppose $A = \text{diag}(\sigma_i)_{i=1}^{n}$, $G \in \mathbb{R}^{r\times n}$, $H \in \mathbb{R}^{n\times r}$, $\text{rank}\, A = n$, $\text{rank}\, G = r(G)$, and $\text{rank}\, H = r(H)$. Then $\text{rank}(GA) = r(G)$, $\text{rank}(AH) = r(H)$, $\sigma_j(GA) \ge \sigma_j(G)\sigma_n$, $\sigma_j(AH) \ge \sigma_j(H)\sigma_n$ for all j.*

Lemma 3.2. *$\sigma_j(GA) = \sigma_j(AH) = \sigma_j(A)$ for all j if G and H are square unitary matrices.*

Lemma 3.3. *[SST06, Proposition 2.2]. Suppose $W \in \mathcal{G}^{m\times n}_{\mu,\sigma}$, $SS^T = S^TS = I_m$, $TT^T = T^TT = I_n$. Then $SW \in \mathcal{G}^{m\times n}_{\mu,\sigma}$ and $WT \in \mathcal{G}^{m\times n}_{\mu,\sigma}$.*

Lemma 3.4. *[SST06, Lemma A.2]. Suppose $\mathbf{a} \in \mathcal{G}^{n\times 1}_{\mu,\sigma}$, $\mathbf{b} \in \mathbb{R}^{n\times 1}$, and $||\mathbf{b}|| = 1$. Then $F_{\mathbf{a}^T\mathbf{b}}(y) \le \sqrt{\frac{2}{\pi}}\frac{y}{\sigma}$.*

Theorem 3.1. *Suppose $G \in \mathbb{R}^{r(G)\times m}$, $H \in \mathbb{R}^{n\times r(H)}$, $\text{rank}\, G = r(G)$, $\text{rank}\, H = r(H)$, $y \ge 0$, and the assumptions of Theorem 2.4 hold. Write $l(G) = \min\{r(G), n\}$ and $l(H) = \min\{m, r(H)\}$. Define $c(r) = 2.35$ for $r > 1$ and $c(1) = \sqrt{\frac{2}{\pi}}$. Then*

(a) $F_{GA}(y) \le c(r(G))y\sqrt{l(G)}/(\sigma_{r(G)}(G)\sigma)$ and

(b) $F_{AH}(y) \le c(r(H))y\sqrt{l(H)}/(\sigma_{r(H)}(H)\sigma)$.

The theorem shows that $\sigma_{\mathrm{rank}(AH)} \le y$ with a probablilty of order at most $1/y$.

Proof. Lemma 3.4 and Fact 2.1 together imply part (a) for $r(G) = 1$ and part (b) for $r(H) = 1$.

To prove part (a) for $r(G) > 1$ assume full SVD $G = S_G \Sigma_G T_G^T$ where $\Sigma_G = (\widehat{\Sigma}_G \mid O)$ and $\widehat{\Sigma}_G = \mathrm{diag}(\sigma_j(G))_{j=1}^{r(G)}$. Write $A_{r(G)} = (I_{r(G)} \mid O) T_G^T A$, $M_{r(G)} = (I_{r(G)} \mid O) T_G^T M$, $W_{r(G)} = (I_{r(G)} \mid O) T_G^T W$, and so $A_{r(G)} = M_{r(G)} + W_{r(G)}$.

We have $GA = S_G \Sigma_G T_G^T A$ and by virtue of Lemma 3.2, $\sigma_j(GA) = \sigma_j(\Sigma_G T_G^T A)$ for all j because S_G is a square unitary matrix. Furthermore $\widehat{\Sigma}_G A_{r(G)} = \Sigma_G T_G^T A$, so $\sigma_j(\Sigma_G T_G^T A) = \sigma_j(\widehat{\Sigma}_G A_{r(G)})$ for all j. By virtue of Lemma 3.1 we have $\sigma_j(\widehat{\Sigma}_G A_{r(G)}) \ge \sigma_{r(G)}(G) \sigma_j(A_{r(G)})$ for all j. We also recall that T_G is a square unitary matrix, apply Lemma 3.3 to the matrix $T_G^T W$, and conclude that this is a Gaussian random matrix having a mean μ and a variance σ^2. Therefore so is its block submatrix $W_{r(G)}$ as well. We can apply Theorem 2.4 for A, M, and W replaced by $A_{r(G)}$, $M_{r(G)}$, and $W_{r(G)}$, respectively, to estimate the CDF $F_{A_{r(G)}}(y)$. By combining this estimate with the above bounds for $j = r(G)$, we prove part (a) of the theorem.

Part (a) implies part (b) because $\sigma_j(AH) = \sigma_j((AH)^T) = \sigma_j(H^T A^T)$ for all j. □

Corollary 3.1. *Under the assumptions of Theorem 3.1 let $m = n$ and $\mu = 0$ and choose two scalars y and z such that $y > 0$ and $z \ge 2\sigma\sqrt{n}$. Then*

(a) $F_{\kappa(GA)}(||G||yz) \ge 2 - \exp(\frac{(z-2\sigma\sqrt{n})^2}{2\sigma^2}) - c(r(G))y\sqrt{l(G)}/(\sigma_{r(G)}(G)\sigma)$ *and*

(b) $F_{\kappa(AH)}(||H||yz) \ge 2 - \exp(\frac{(z-2\sigma\sqrt{n})^2}{2\sigma^2}) - c(r(H))y\sqrt{l(H)}/(\sigma_{r(H)}(H)\sigma)$.

Proof. Combine Theorems 2.5 for $y = z$ and 3.1. □

The following corollary extends the lower bounds of Theorem 3.1 for a randomized matrix product to the respective bounds for its leading blocks; this implies that *randomized multiplication is expected to be generic preconditioning.*

Corollary 3.2. *Suppose $j, k, m,, n$, q and s are integers, $1 \le j \le q$, $1 \le k \le s$, $G \in \mathbb{R}^{q \times m}$, $H \in \mathbb{R}^{n \times s}$, $A = M + W$, $W \in \mathcal{G}_{\mu,\sigma}^{m \times n}$, and $y \ge 0$. Write $G_j = (I_j \mid O_{j,m-j})G$, $r(G_j) = \mathrm{rank}(G_j)$ and $l(G_j) = \min\{r(G_j), n\}$ for $j = 1, \dots, q$, $H_k = H \begin{pmatrix} I_k \\ O_{n-k,k} \end{pmatrix}$, $r(H_k) = \mathrm{rank}(H_k)$ and $l(H_k) = \min\{m, r(H_k)\}$ for $k = 1, \dots, s$. Then for all j and k in the above ranges we have*

(a) $F_{(GA)^{(j)}}(y) \le c(r(G))y\sqrt{l(G_j)}/(\sigma_{r(G_j)}\sigma)$ *and*

(c) $F_{(AH)^{(k)}}(y) \le c(r(H))y\sqrt{l(H_k)}/(\sigma_{r(H_k)}\sigma)$.

Proof. For every j and every k apply Theorem 3.1 replacing G by G_j, H by H_k, and A by either $A \begin{pmatrix} I_j \\ O_{n-j,j} \end{pmatrix}$ or $(I_k \mid O_{k,m-k})A$. □

Corollary 3.3. *Under the assumptions of Corollary 3.2 let $\sigma \ne 0$. Then with probability one the matrices GA and AH have generic rank profile.*

Combining the latter results with Theorem 2.1 implies that randomized multiplication is expected to make Gaussian elimination with no pivoting numerically safe, and similarly for block Gaussian elimination (cf. [PQZa]).

Corollary 3.4. *Suppose M is a normalized $m \times n$ well conditioned matrix of full rank, $||M|| = 1$, $B \in \mathcal{G}_{0,1}^{m \times m}$ and $C \in \mathcal{G}_{0,1}^{n \times n}$. Then Gaussian elimination with no pivoting applied to computing triangular factorizations of the matrices BM and MC is expected to involve no divisions by absolutely small values.*

4 Randomized additive preconditioning

Suppose a matrix $A \in \mathbb{R}^{n\times n}$ has a numerical nullity r, $0 < r < n$, $U, V \in \mathcal{G}_{0,\sigma}^{n\times r}$ for $\sigma = ||A||/(2\sqrt{r})$, and $C = A + UV^T$. Our goal is to employ the results of the previous section to prove that additive preprocessing $A \Longrightarrow C = A+UV^T$ is expected to improve conditioning of the ill conditioned matrix A (cf. Remark 4.2).

We first reduce the study of ill conditioned, that is nearly singular input matrix A to the case of a nearby singular matrix $\tilde{A}$ and then use its SVD and factorizations of some auxiliary matrices.

Proceeding orderly, let $\tilde{A} = A + E$ be the matrix of a rank $\rho = n - r$ obtained by zeroing the singular values $\sigma_j(A)$ for $j > n - r$ in the SVD $A = S\Sigma T^T$, so that $||E|| = \sigma_{n-r+1}(A)$.

Write $C = A + UV^T$ and $\tilde{C} = \tilde{A} + UV^T = C + E$, assume that the matrices C and $\tilde{C}$ are nonsingular and recall that $\kappa(\tilde{C}) \le \frac{1+\delta}{1-\delta\kappa(C)}\kappa(C)$ where $\delta = \frac{||E||}{||C||}$ and $\delta\kappa(C) < 1$ [GL96, Section 5.5.5].

Next assume that the value δ is small, write $\tilde{C} = \tilde{A} + UV^T$, and in the rest of this section estimate the ratio $\frac{\kappa(\tilde{C})}{\kappa(\tilde{A})}$, which closely approximates the ratio $\frac{\kappa(C)}{\kappa(A)}$.

Furthermore, until arriving at Corollary 4.3 we simplify the notation by dropping the character "tilde" and writing A and C instead of $\tilde{A}$ and $\tilde{C}$ assuming that $\operatorname{rank} A = n - r$ and $C = A + UV^T$.

The following results are readily verified.

Theorem 4.1. *Let $A = S\Sigma T^T$ be full SVD of an $n \times n$ matrix A of a rank ρ where $\rho < n$, S and T are unitary matrices, S, $T \in \mathbb{C}^{n\times n}$, $\Sigma = \operatorname{diag}(\Sigma_A, O_{r,r})$ is an $n \times n$ diagonal matrix, $r = n - \rho$, and $\Sigma_A = \operatorname{diag}(\sigma_j)_{j=1}^{\rho}$ is the $\rho \times \rho$ diagonal matrix of the positive singular values of the matrix A. Suppose $U \in \mathbb{C}^{n\times r}$, $V \in \mathbb{C}^{n\times r}$, and let the $n \times n$ matrix $C = A + UV^T$ be nonsingular. Write*

$$S^TU = \begin{pmatrix} U_\rho \\ U_r \end{pmatrix},\ T^TV = \begin{pmatrix} V_\rho \\ V_r \end{pmatrix},\ R_U = \begin{pmatrix} I_\rho & U_\rho \\ O & U_r \end{pmatrix},\ R_V = \begin{pmatrix} I_\rho & V_\rho \\ O & V_r \end{pmatrix}$$

where U_r and V_r are nonsingular $r \times r$ matrices. Then $R_U\Sigma R_V^T = \Sigma$, $R_U \operatorname{diag}(O_{\rho,\rho}, I_r)R_V^T = S^TUV^TT$, so that

$$C = SR_U \operatorname{diag}(\Sigma_A, I_r)R_V^TT^T. \tag{4.1}$$

Theorem 4.2. *Under the assumptions of Theorem 4.1, write $p = ||R_U^{-1}||\ ||R_V^{-1}||$ and let $||A|| = 1$. Then (a) $p \ge \frac{||C^{-1}||}{||A^+||} = \frac{\sigma_{n-r}(A)}{\sigma_n(C)}$ and* (b) $1 \le p^2 \le (1 + (1 + ||U||^2)||U_r^{-1}||^2)(1 + (1 + ||V||^2)||V_r^{-1}||^2)$.

Proof. (a) Invert matrix equation (4.1) and obtain that $C^{-1} = TR_V^{-T} \operatorname{diag}(\Sigma_A^{-1}, I_r)R_U^{-1}S^T$. Therefore $||C^{-1}|| = ||R_V^{-T} \operatorname{diag}(\Sigma_A^{-1}, I_r)R_U^{-1}|| \le ||R_V^{-T}||\ ||\operatorname{diag}(\Sigma_A^{-1}, I_r)||\ ||R_U^{-1}||$ because S and T are square unitary matrices. Recall that $\sigma_{n-r}(A) \le ||A|| = 1$ by assumption, substitute the expression $||\operatorname{diag}(\Sigma_A^{-1}, I_r)|| = 1/\sigma_{n-r}(A) = ||A^+||$, and obtain the claimed upper bound p on the ratio $\frac{||C^{-1}||}{||A^+||}$.

(b) Combine the equations $R_U^{-1} = \begin{pmatrix} I_\rho & -U_\rho U_r^{-1} \\ O & U_r^{-1} \end{pmatrix}$ and $R_V^{-1} = \begin{pmatrix} I_\rho & -V_\rho V_r^{-1} \\ O & V_r^{-1} \end{pmatrix}$. □

Now we normalize the matrix A by scaling to ensure that $||A|| = 1$, let $\sigma = ||A||/(2\sqrt{r})$, and choose two matrices $U, V \in \mathcal{G}_{0,\sigma}^{n\times r}$. Theorem 2.5 implies that $F_{||W||}(y) \ge 1 - \exp(-2(y-1)^2/r)$ provided $y \ge 1$ and $W = U$ or $W = V$. We arrive at the expected bounds $||U|| \le 1$, $||U|| \le 1$, $||UV^T|| \le ||U||\ ||V|| \le 1$, $||C|| \le 2$ and $p \le 1 + 2f_r$ for $f_r = \max\{||U_r^{-1}||^2, ||V_r^{-1}||^2\}$ where the probability that the norm $||U||$ or $||V||$ exceeds one by at least δ decreases to zero exponentially in $2\delta^2/r$. This implies expenentially rapid decrease of the probability that $||C|| \ge 2+\delta$ or $p \ge 1+2f_r+\delta$ as $\delta \to \infty$.

Next we bound the values $||U_r^{-1}|| = \frac{1}{\sigma_r(U_r)}$ and $||V_r^{-1}|| = \frac{1}{\sigma_r(V_r)}$ based on Theorem 4.2.

Theorem 4.3. *Let $U, V \in \mathcal{G}_{\mu,\sigma}^{n\times r}$, U_r, and V_r denote the four matrices in Theorem 4.1, suppose $m = n$ and Theorem 3.1 holds*

(a) for $r(G) = r$, $G = (O \mid I_r)S^T$, $A = U$ (in this case $GA = U_r$) as well as

(b) for $r(G) = r$, $m = n$, $G = (O \mid I_r)T^T$, $A = V$ (in this case $GA = V_r$).

Also assume $F_A(y)$ in Definition 2.1 and $c(r)$ in Theorem 3.1, $c(r) = 2.35$ for $r > 1$ and $c(1) = \sqrt{\frac{2}{\pi}}$. Then

(a) $F_{U_r}(y) \le c(r)\ y\sqrt{r}/\sigma$ and (b) $F_{V_r}(y) \le c(r)\ y\sqrt{r}/\sigma$, respectively.

Proof. Apply part (a) of Theorem 3.1 for $r(G) = r$, $G = (O, I_r)S^T$ and $A = U$ to obtain that $F_{U_r}(y) \le c(r)y\sqrt{r}/(\sigma_r((O \mid I_r)S^T)\sigma)$. Then apply part (a) of Theorem 3.1 for $r(G) = r$, $m = n$, $G = (O \mid I_r)T^T$ and $A = V$ to obtain that $F_{V_r}(y) \le c(r)y\sqrt{r}/(\sigma_r((O, I_r)T^T)\sigma)$. Observe that $\sigma_r((O \mid I_r)S^T) = \sigma_r((O \mid I_r)T^T) = 1$ because S and T are unitary matrices. Substitute these equations into the above bounds on $F_{U_r}(y)$ and $F_{V_r}(y)$, and obtain both parts (a) and (b) of Theorem 4.3. □

Corollary 4.1. *Under the assumptions of Theorem 4.3 we have*

(a) the matrix C is singular with probability zero and

(b) if in addition $||A|| = 1$, $\mu = 0$ and $\sigma = \frac{1}{2\sqrt{r}}$, then Probability$\{\frac{\kappa(C)}{\kappa(A)} \ge 2\frac{\sigma_1(A)}{\sigma_{n-r}(A)} + z^2\} = O(1/z)$ *as $z \to \infty$.*

Proof. Theorem 4.3 implies that the matrices U_r and V_r are singular with probability zero. Therefore part (a) of the corollary follows from equation (4.1). We deduce part (b) by combining Theorem 4.2 with our probabilistic upper bounds on the norms $||U||$, $||V||$, and $||C||$, and the probabilistic upper bounds of Theorem 4.3 on the norms $||U_r^{-1}||$ and $||V_r^{-1}||$. □

Corollary 4.2. *Under the assumptions of Corollary 4.1 the matrix C is expected to be nonsingular with probability one and to have condition number of order at most $\sigma_1(A)/\sigma_{n-r}(A)$; the probability of exceeding this bound is inversely proportional to the square root of the excess value.*

Corollary 4.3. *Keep the assumptions of Corollary 4.1, but allow the matrix A to have full rank n and numerical rank $n - r$, so that the ratio $\sigma_1(A)/\sigma_{n-r}(A)$ is not large. Then still the matrix C is expected to be nonsingular with probability one and to have condition number of order at most $\sigma_1(A)/\sigma_{n-r}(A)$; the probability of exceeding this bound is inversely proportional to the square root of the excess value (cf. Section 5.3).*

Proof. Obtain the matrix $\tilde{A} = A + E$ of rank $n - r$ by setting the r smallest singular values of the matrix A to zero. They are assumed to be small, and so the norm $||E||$ is small. Apply Corollary 4.2 to that matrix and deduce that the matrix $\tilde{C} = \tilde{A} + UV^T$ is expected to be well conditioned. Consequently (cf. Theorem 2.2) the transition back to the matrix $C = \tilde{C} - E$ keeps the same properties stated in Corollary 4.2 because the norm $||E||$ is small. □

Remark 4.1. *One can readily deduce from Corollary 2.4 that the matrix C is likely to be nonsingular also where the entries of the matrices U and V have been sampled randomly and uniformly from a set of a large cardinality in any ring.*

Remark 4.2. *In part (b) of Corollary 4.1 we assume that $\mu = 0$ to simplify our proofs, but the result of Corollaries 4.2 and 4.3 can be extended to any choice of a mean μ of order $||A||$ or less. Likewise one can assume more relaxed lower and upper bounds on the ratio $||A||/\sigma$.*

Remark 4.3. *We can preserve the regularization and preconditioning power of the map $A \Longrightarrow C$ even where we choose $U = aV$ for a fixed nonzero scalar a and thus use fewer random parameters.*

Remark 4.4. *Our map $A \Longrightarrow C$ is expected to produce a well conditioned matrix C provided that the ratio $\frac{\sigma_1(A)}{\sigma_{n-r}(A)}$ is not large, but such a ratio is large only for matrices A lying near the algebraic variety of matrices M that have ranks at most $n - r - 1$, that is such that $\sigma_{n-r}(M) = 0$. Such a variety is empty where $r = n-1$ and generally has dimension $n^2-(r+1)^2$. Indeed let $M = \begin{pmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{pmatrix}$ where the leading block M_{00} is a nonsingular $(n - r - 1) \times (n - r - 1)$ matrix. Then $M_{11} = M_{10}M_{00}^{-1}M_{01}$, and the claimed bound on the dimension follows because similar argument can be applied in the case where any $(n - r - 1) \times (n - r - 1)$ submatrix of the matrix M is nonsingular.*

Remark 4.5. *Assume the dual additive preprocessing* $A^{-1} \Longrightarrow (C_-)^{-1} = A^{-1} + U_- V_-^T$ *of (2.9), where* $C_- = A - AU_- H^{-1} V_-^T A$, $U_-, V_- \in \mathbb{C}^{n\times q}$, *and* $H = I_q + V_-^T A U_-$. *Then our analysis implies that the condition number* $\kappa(C)_-$ *is expected to have order* $\sigma_{q+1}(A)/\sigma_n(A)$ *provided* U_- *and* V_- *are standard Gaussian random matrices and the matrix* A *has been scaled so that the norm* $||A^{-1}||$ *is neither large nor small (cf. Section 5.3). To define such scaling we need a crude estimate for the norm* $||A^{-1}||$*; we can obtain it at a low computational cost, e.g., by applying the randomized algorithm in [D83].*

Remark 4.6. *Can we extend our study in this section to preprocessing with structured matrices* U *and* V*, defined by fewer than* $2rn$ *parameters? We cannot extend the proof of Theorem 3.1 and consequently of Theorem 4.3, but according to our extensive tests random Toeplitz matrices tend to be well conditioned similarly to random general matrices (see Section 7).*

5 Applications, derandomization and extensions of randomized additive preprocessing

5.1 Application to multiplicative preprocessing

Suppose a real $n \times n$ matrix A has a small positive numerical nullity r, $\sigma/||A|| \approx 1$, $U, V \in \mathcal{G}_{0,\sigma}^{n\times r}$ and $C = A + UV^H$. Then by virtue of Corollary 2.2 the matrix C^{-1} is a level r multiplicative preprocessor for the matrix A, whereas by virtue of Corollary 4.3 the matrix C is expected to be well conditioned and therefore is more readily invertible than the ill conditioned matrix A.

Likewise additive preprocessing enables us to devise a level q multiplicative preprocessor for a nonsingular matrix A having a small numerical rank q. Such a preprocessor is given by the matrix C_-^{-1} where $C_- = A - AU_- H^{-1} V_-^T A$, $H = I_q + V_-^T A U_-$ (cf. equation (2.9) and Remark 4.5) and $U_-, V_- \in \mathcal{G}_{0,\sigma}^{n\times q}$. Indeed $C_-^{-1} = A^{-1} + U_- V_-^T$, and so the matrices $I - C_-^{-1}A$ and $I - AC_-^{-1}$ have rank at most q, whereas the matrix C_- is expected to be well conditioned (assuming that the ratio $\sigma/||A^{-1}||$ is properly bounded from above and below) and therefore is more readily invertible than the matrix A.

5.2 Computation of nmbs, approximation of leading and trailing singular spaces, and approximation by matrices of a small rank and by structured matrices

Our next theorem employs additive preprocessing to compute null vectors of an $n\times n$ singular matrix $\tilde{A}$ with a nullity r and matrix bases for its null space as well as its left null space. Theorem 5.2 extends these results to the approximation of the right and left trailing singular spaces $\mathbb{T}_{A,r}$ and $\mathbb{S}_{A,r}$ as well as the right and left leading singular spaces $\mathbb{T}_A^{(n-r)}$ and $\mathbb{S}_A^{(n-r)}$ of an $n \times n$ nonsingular matrix $A \approx \tilde{A}$ that has a positive numerical nullity r (cf. Section 2.4).

Theorem 5.1. *[PQ10, Theorem 3.1]. Assume that* $\tilde{C} = \tilde{A} + UV^H$, $\tilde{A}, \tilde{C} \in \mathbb{C}^{n\times n}$, $n > r > 0$, $r = \operatorname{nul} \tilde{A} = n - \operatorname{rank} \tilde{A}$, $U, V \in \mathbb{C}^{n\times r}$, *and the matrix* $\tilde{C}$ *is nonsingular. Then the matrix* $\tilde{C}^{-1}U$ *is a* $\operatorname{nmb}(\tilde{A})$*, whereas the matrix* $\tilde{C}^{-T}V$ *is a left* $\operatorname{nmb}(\tilde{A})$.

Theorem 5.2. *(a) Suppose an* $n \times n$ *nonsingular matrix* A *has a numerical nullity* r *and has numerical rank* $q = n - r$, *that is the ratio* $\sigma_1(A)/\sigma_{q+1}(A)$ *is large, but the ratio* $\sigma_1(A)/\sigma_q(A)$ *is not large. Suppose* $n > r > 0$, U *and* V *are* $n \times r$ *matrices, and the matrix* $C = A + UV^H$ *is nonsingular. Then there exist two matrices* B_U *and* B_V *and a scalar* c *independent of* A, U, V, n *and* r *such that* $\mathcal{R}(B_U) = \mathbb{T}_{A,r}$, $\mathcal{R}(B_V) = \mathbb{S}_{A,r}$, $||C^{-1}U - B_U|| \le c\sigma_{q+1}(A)||B_U||$, *and* $||C^{-T}V - B_V|| \le c\sigma_{q+1}(A)||B_V||$.

(b) Furthermore suppose U_- *and* V_- *are* $n \times q$ *matrices for* $q = n - r$, *write* $H = I_q + V_- A U_-^H$ *and* $C_- = A - AU_- H^{-1} V_-^H A$, *and assume that the matrix* H *is nonsingular. Then there exist two matrices* B_{U_-} *and* B_{V_-} *and a scalar* c_- *independent of* A, U_-, V_-, n *and* q *such that* $\mathcal{R}(B_{U_-}) = \mathbb{T}_A^{(q)}$, $\mathcal{R}(B_{V_-}) = \mathbb{S}_A^{(q)}$, $||C_- U_- - B_{U_-}|| \le c_- \sigma_{r+1}(A)||B_{U_-}||$, *and* $||C_-^H V_- - B_{V_-}|| \le c_- \sigma_{r+1}(A)||B_{V_-}||$.

Proof. See [PQ10, Section 7.1]. □

[PQ10, Theorem 3.1 and Section 7.1] also cover the more general case of rectangular matrices A.

Part (a) of Theorem 5.2 shows that $\mathcal{R}(C^{-1}U) \approx \mathbb{T}_{A,r}$ and $\mathcal{R}(C^{-T}V) \approx \mathbb{S}_{A,r}$, that is, the linear spaces $\mathcal{R}(C^{-1}U)$ and $\mathcal{R}(C^{-T}V)$ approximate the right and left trailing singular spaces associated with the r smallest singular values of the matrix A, respectively. Likewise part (b) shows that $\mathcal{R}(C_-U_-) \approx \mathbb{T}_A^{(q)}$ and $\mathcal{R}(C_-^HV_-) \approx \mathbb{S}_A^{(q)}$), that is, the linear spaces $\mathcal{R}(C_-U_-)$ and $\mathcal{R}(C_-^HV_-)$ approximate the right and left leading singular spaces associated with the q largest singular values of the matrix A, respectively.

If a matrix A has a small numerical rank r we can closely approximate it by the matrix AQQ^H of rank r where $Q = Q(C_-U_-)$ and the matrix C_- can be readily computed based on the dual SMW formula (2.9). This computation is division-free except for orthogonalization of the $n \times r$ matrix C_-U_- and the inversion of the $r \times r$ matrix H, which we can make diagonally dominant (and therefore nonsingular) by properly scaling the matrices A, U and V.

We refer the reader to [HMT11] on numerous highly important applications of the approximation by matrices of a small ranks and note that our algorithm for this approximation is quite competitive, e.g. we do not need to involve auxiliary matrices of larger sizes than $n \times r$ and by making the matrix H nonsingular we ensure that our randomization does not fail.

By applying this approximation algorithm to the displacement of a matrix lying near a matrix with displacement structure we obtain its approximation by a matrix having a small displacement rank. Such approximations are involved, e.g. into Newton's structured matrix inversion (see the Appendix).

A natural and important extension to low-rank decompositions of tensors is the subject of high research interest (cf. [MMD08], [OT09]).

5.3 Derandomization of additive preprocessing

Let us show that a small amount of additional computations enables derandomization of our additive preprocessing (cf. [PQ10, Theorem 8.1]).

Theorem 5.3. *Assume matrices $A \in C^{n\times n}$, unitary $U, V, U_-, V_- \in C^{n\times r}$ and $C = A + UV^H$ such that $0 < r < n$, $q = n - r$, $\sigma_1(A) \geq 1 \geq \sigma_q(A)$, $\mathrm{rank}(UV^H) = r$ and $\mathrm{rank}(C) = n$. Write $U_1 = Q(C^{-1}U)$, $V_1 = Q(C^{-H}V)$, and $C_1 = A + U_1V_1^H$.*

(a) If the matrix A has rank q, then the matrix C_1 is nonsingular and $\kappa(C_1) = \kappa(A) = \sigma_1(A)/\sigma_q(A)$.

(b) If the matrix A has numerical rank q, then the matrix C_1 is nonsingular and $\kappa(C_1) \approx \sigma_1(A)/\sigma_q(A)$.

(c) (cf. Remark 4.5). Suppose the matrices A and $C_-^{-1} = A^{-1} + U_-V_-^H$ are nonsingular and the matrix A has numerical rank $q = n - r$. Write $U_-^{(1)} = Q(C_-U_-)$, $V_-^{(1)} = Q(V_-^HC_-)$, $H = I_q + V_-^{(1)}A(U_-^{(1)})^H$ and $C_-^{(1)} = A - AU_-^{(1)}H^{-1}(V_-^{(1)})^HA$. Then the matrix $C_-^{(1)}$ is nonsingular, $(C_-^{(1)})^{-1} = A^{-1} + U_-^{(1)}(U_-^{(1)})^H$, and $\kappa(C_-^{(1)}) \approx \sigma_{q+1}(A)/\sigma_n(A)$.

Proof. Due to Theorem 5.1, the updated matrices U_1 and V_1 remain the right and left nmbs for the matrix A, respectively. Let $A = \sum_{j=1}^q \sigma_j \mathbf{s}_j\mathbf{t}_j^H$ be an SVD of the matrix A. Write $U_1 = (\mathbf{u}_j)_{j=1}^r$ and $V_1 = (\mathbf{v}_j)_{j=1}^r$ and obtain the SVD of the matrix $C_1 = A + U_1V_1^H = \sum_{j=1}^r \mathbf{u}_j\mathbf{v}_j^H + \sum_{j=1}^q \sigma_j\mathbf{s}_j\mathbf{t}_j^H$. This implies part (a) of the theorem because $r = n - q$ and $\sigma_1 \geq 1 \geq \sigma_q$. Parts (b) follows from part (a) by the continuity argument. Apply part (b) to matrices A^{-1}, U_- and V_- replacing A, U and V, respectively, and obtain part (c). □

The theorem derandomizes the estimates of Corollary 4.3 and Remark 4.5. Over the real or complex numbers our derandomization can only fail with probability zero, namely where the matrix $A + UV^H$ in parts (a) and (b) or $A^{-1} + U_-V_-^H$ in part (c) is singular.

5.4 Solution of a linear system of equations with preconditioning based on computation of nmbs and approximation of trailing singular spaces

Theorem 5.1 reduces the computation of null vectors and nmbs to the solution of some nonsingular linear systems of equations. Conversely the solution of a nonsingular linear system of n equations $A\mathbf{y} = \mathbf{b}$ can be expressed via the null vector $\begin{pmatrix} \mathbf{y} \\ -1/\beta \end{pmatrix}$ of the matrix $K = (A \mid \beta\mathbf{b})$ for a nonzero scalar β. If the matrix A has numerical nullity one and if the ratio $||A||/||\beta\mathbf{b}||$ is neither large nor small, then on the average vector $\mathbf{b}$ the map $A \Longrightarrow K$ serves as preconditioning [PQa].

Our two next alternative preconditioning algorithms employ randomized pre- and post-multiplication based on parts (a) and (b) of Theorem 5.2, respectively, to define approximate 2×2 block Gauss–Jordan diagonalization of an input matrix. One can employ pre-multiplication alone (that is remove the post-multipliers $(L_0 \mid L_1)$ from Algorithms 5.1 and 5.2); then one would expect to arrive at an approximation to a 2×2 block triangular matrix.

Assume that the $n \times n$ nonsingular input matrix A has a small positive numerical nullity r (that is assume that the ratio $\frac{\sigma_1(A)}{\sigma_{n-r}(A)}$ is not large, whereas $\sigma_1(A) \gg \sigma_{n-r+1}(A)$) and devise the following randomized preconditioning algorithm.

Algorithm 5.1. Preconditioning based on the approximation of trailing singular spaces.

INPUT: *Two integers n and r, $0 < r < n$, a nonsingular matrix $A \in \mathbb{C}^{n\times n}$ having numerical rank $n-r$ and scaled so that $||A|| = 1$, and a Subroutine* LIN·SOLVE *that solves a linear system of equations if it is nonsingular and well condtioned or outputs FAILURE otherwise.*

OUTPUT: *FAILURE or four matrices K_0 and L_0 in $\mathbb{C}^{n\times(n-r)}$ and K_1 and L_1 in $\mathbb{C}^{n\times r}$ such that $W = (K_0 \mid K_1)^H A(L_0 \mid L_1) = \begin{pmatrix} W_{00} & W_{01} \\ W_{10} & W_{11} \end{pmatrix}$ and with a probability near one the block submatrix $W_{00} = K_0^H A L_0$ is nonsingular, well conditioned, and strongly dominant, that is $||W_{00}|| \gg \max\{||W_{01}||,\ ||W_{10}||,\ ||W_{11}||\}$.*

INITIALIZATION: *Generate four standard Gaussian random matrices S and T in $\mathbb{C}^{n\times(n-r)}$, U and V in $\mathbb{C}^{n\times r}$.*

COMPUTATIONS:

1. *Compute the matrix $C = A + UV^H$ (expected to be nonsingular and well condtioned according to our study in Section 4).*
2. *Apply the Subroutine* LIN·SOLVE *to compute and to output the matrices $K_1 = C^{-T}V$ and $L_1 = C^{-1}U$. Stop and output FAILURE if so does the subroutine.*
3. *Output the matrices $K_0 = S$ and $L_0 = T$ and stop.*

Correctness of the algorithm follows because the value $\sigma_{n-r}(W_{00})$ is likely to have order $\sigma_{n-r}(A)$ by virtue of Theorem 3.1, whereas the matrices W_{01}, W_{10}, and W_{11} have the norms of at most order $\sigma_{n-r+1}(A)$. This is because, by virtue of part (a) of Theorem 5.2, the matrices K_1 and L_1 are approximate trailing singular spaces of the matrices A^H and A, respectively, and because the matrix A has numerical nullity r, which implies that $\sigma_{n-r}(A) \gg \sigma_{n-r+1}(A)$, whereas the ratio $\sigma_1(A)/\sigma_{n-r}(A)$ is not large.

Since the block W_{00} in the 2×2 block matrix W is dominant, nonsingular and well conditioned, we can apply block Gauss–Jordan diagonalization and readily factorize the matrix W as follows,

$$W = \begin{pmatrix} I & O \\ W_{10}W_{00}^{-1} & I \end{pmatrix} \begin{pmatrix} W_{00} & O \\ O & G \end{pmatrix} \begin{pmatrix} I & W_{00}^{-1}W_{01} \\ O & I \end{pmatrix}$$

where $G = W_{11} - W_{10}W_{00}^{-1}W_{01}$.

Based on this factorization, one can immediately reduce the inversion of the matrices W and A and the solution of a linear system $A\mathbf{y} = \mathbf{b}$ to the similar operations with the matrices W_{00} and G of smaller sizes, expected to be nonsingular and better conditioned (cf. (2.8) and Corollary 4.3).

Tables 7.8 and 7.9 demonstrate substantial advantages of this approach versus standard algorithms for solving linear systems of equations.

Remark 5.1. *Computation of the Schur complement G involves $O(n^2r)$ flops. This r/n fraction of all flops used should be performed in extended precision to counter the expected cancellation of the leading digits of the input values. One can orthogonalize the matrices K_0 and L_0 and apply derandomization in Theorem 5.3, part (b), to stabilize these computations numerically.*

Remark 5.2. *To work with fewer random parameters one can generate a single $n \times q$ standard Gaussian random matrix for $q \geq \max\{r, n-r\}$ and then reuse its columns while defining the four auxiliary random matrices U, V, K_0, and L_0.*

5.5 Preconditioning based on approximation of leading singular spaces

In the case of matrices A having a small numerical rank, one can reduce the solution of a linear system $A\mathbf{y} = \mathbf{b}$ to the approximation of leading singular spaces based on part (b) of Theorem 5.2, which employs dual additive preprocessing and leads to dual version of Algorithm 5.1. In this version a crude approximation to the norm A^{-1} of an ill conditioned input matrix A is required (see [D83] on fast randomized computation of such an approximation), but neither matrix inversion nor solving linear systems of equations are involved, except for the inversion of an auxiliary $q \times q$ matrix H, expected to be close to the identity matrix I_q.

Then again we specify our algorithm for approximate 2×2 block Gauss–Jordan diagonalization of an input matrix. One can employ pre-multiplication alone (by removing the multiplier $(L_0 \mid L_1)$ from Algorithm 5.2) and then would expect to arrive at an approximation to a 2×2 block triangular matrix.

Algorithm 5.2. Dual preconditioning based on the approximation of leading singular spaces.

INPUT: *A Subroutine* LIN·SOLVE *that solves a linear system of equations if it is nonsingular and well conditioned or outputs FAILURE otherwise, two integers n and q, $0 < q < n$, and a nonsingular matrix $A \in \mathbb{C}^{n \times n}$ having numerical rank q and scaled so that $||A^{-1}|| = 1$, which implies that the norm $||A||$ is small since the matrix A is ill conditioned under the above assumption.*

OUTPUT: *FAILURE or four matrices K_0, $L_0 \in \mathbb{C}^{n \times q}$ and K_1, $L_1 \in \mathbb{C}^{n \times (n-q)}$ such that $W = (K_0 \mid K_1)^H A (L_0 \mid L_1) = \begin{pmatrix} W_{00} & W_{01} \\ W_{10} & W_{11} \end{pmatrix}$ and with a high probability the block submatrix $W_{00} = K_0^H A L_0$ is nonsingular, well conditioned, and strongly dominant, that is*

$$||W_{00}|| \gg \max\{||W_{01}||,\ ||W_{10}||,\ ||W_{11}||\}.$$

INITIALIZATION: *Generate four standard Gaussian random matrices S and T in $\mathbb{C}^{n \times (n-q)}$ and U_- and V_- in $\mathbb{C}^{n \times q}$.*

COMPUTATIONS:

1. *Compute the matrix $H = I_q + V_- A U_-^H$, expected to be close to the identity matrix I_q since the norm $||A||$ is small.*
2. *Apply the Subroutine* LIN·SOLVE *to compute the matrix H^{-1}. Stop and output FAILURE if so does the subroutine.*
3. *Compute the matrix $C_- = A - A U_- H^{-1} V_-^H A$.*
4. *Compute and output the matrices $K_1 = C_-^H V_-$ and $L_1 = C_- U_-$.*

5. Compute and output the matrices $K_0 = S$ and $L_0 = T$ and stop.

Both Algorithms 5.1 and 5.2 can be extended to the block factorization of rectangular inputs A. Remarks 5.1 and 5.2 can be readily extended as well. We only specify an extension of Remark 5.1.

Remark 5.3. *Computation of the auxiliary matrix H involves $O(n^2q)$ flops. This q/n fraction of all flops used should be performed in extended precision to counter the expected cancellation of the leading digits of the input values. One can orthogonalize the matrices K_0 and L_0 and apply derandomization in Theorem 5.3, part (c), to stabilize these computations numerically.*

5.6 Applications of randomized augmentation

Based on our next theorem we can extend the properties of additive preprocessing to the southeastern augmentation $A \Longrightarrow K = \begin{pmatrix} A & -U \\ WV^H & W \end{pmatrix}$ where $W \in \mathbb{C}^{r\times r}$, $U, V \in \mathbb{C}^{n\times r}$, $K \in \mathbb{C}^{(n+r)\times(n+r)}$, the matrix A is nonsingular, U, V, and W are random scaled matrices, and $0 < r < n$. Randomized augmentation can serve as an alternative to randomized additive preprocessing because both techniques are closely linked to one another, e.g. via the following simple factorizations.

Theorem 5.4. *Suppose $K = \begin{pmatrix} A & -U \\ WV^H & W \end{pmatrix} \in \mathbb{C}^{(m+r)\times(n+r)}$, $W \in \mathbb{C}^{r\times r}$ is a nonsingular matrix, $C = A + UV^H$. Then $K = \operatorname{diag}(I_m, W)\widehat{U}\operatorname{diag}(C, I_r)\widehat{V}$ for $\widehat{U} = \begin{pmatrix} I_m & -U \\ O_{r,m} & I_r \end{pmatrix}$, $\bar{U} = \widehat{U}^{-1} = \begin{pmatrix} I_m & U \\ O_{r,m} & I_r \end{pmatrix}$, $\widehat{V} = \begin{pmatrix} I_n & O_{n,r} \\ V^H & I_r \end{pmatrix}$, $\bar{V} = \widehat{V}^{-1} = \begin{pmatrix} I_n & O_{n,r} \\ -V^H & I_r \end{pmatrix}$. Moreover if the matrices C and K are square and nonsingular, then we have $K^{-1} = \bar{V}\operatorname{diag}(C^{-1}, I_r)\bar{U}\operatorname{diag}(I_m, W^{-1})$ and consequently $C^{-1} = (I_n \mid O_{n,r})K^{-1}(I_m \mid O_{m,r})^T$.*

One can similarly employ the northwestern augmentation

$$A \Longrightarrow \begin{pmatrix} W & WV^H \\ -U & A \end{pmatrix} = \begin{pmatrix} O_{r,m} & I_r \\ I_m & O_{m,r} \end{pmatrix} K \begin{pmatrix} O_{n,r} & I_n \\ I_r & O_{r,n} \end{pmatrix} \tag{5.1}$$

as well as northeastern and southwestern augmentations.

In [PQa] the preconditioning property is proved based on Theorem 3.1 for the more general class of augmentations $K = \begin{pmatrix} A & -U \\ V^H & W \end{pmatrix}$. Namely we can expect to have $\kappa(K)$ of order at most $\sigma_1(A)/\sigma_{l-r}(A)$ for $l = \min\{m, n\}$ provided $U \in \mathcal{G}_{0,\sigma}^{m\times r}$, $V \in \mathcal{G}_{0,\sigma}^{n\times r}$, and $W \in \mathcal{G}_{0,\sigma}^{r\times r}$ are three random matrices and $\sigma \approx ||A||$. Based on this property and on Theorem 5.4 we can apply randomized augmentation similarly to randomized additive preprocessing. We can devise and analyze the respective algorithms directly, without reduction of augmentation to additive preprocessing. Here is an example from [PQa, Section 3.1].

Theorem 5.5. *Assume two matrices $A \in \mathbb{C}^{m\times n}$ of a rank $\rho < n$ and $V \in \mathbb{C}^{r\times n}$ for $r = n - \rho$. Suppose the matrix $C = \begin{pmatrix} V \\ A \end{pmatrix}$ has full column rank n. Then $B = C^{(I)}\begin{pmatrix} I_r \\ O \end{pmatrix}$ is an $nmb(A)$.*

Augmentation enables us to preserve matrix structure (e.g. of Toeplitz and Hankel types) substantially better than additive preprocessing can do this (see Section 6).

Remark 5.4. *(Cf. Remark 5.4.) As in the case of additive preprocessing, we can preserve the regularization and preconditioning power of the augmentation map $A \Longrightarrow K$ even where we choose $U = aV$ for a fixed nonzero scalar a. In this case we need fewer random parameters. We should avoid the augmentations $A \Longrightarrow K = \begin{pmatrix} A & U \\ V^H & W \end{pmatrix}$ producing a Hermitian positive definite matrices K. Otherwise the augmentation would not decrease the condition number due to the Interlacing Property of the eigenvalues of Hermitian matrices [GL96, Theorem 8.1.7]. We do not have such a problem with additive preprocessing (cf. [W07]).*

Remark 5.5. *One can embed an input matrix A into a larger matrix banded with zeros and then view augmentation as its $2r$-rank perturbation. One can apply such a perturbation to the matrix A itself (we call this semi-augmentation). E.g., one can replace the entries in the northeastern or southwestern corner of a Toeplitz matrix A by random values but preserve its Toeplitz structure.*

5.7 Randomized additive preconditioning with the SMW recovery

Suppose we seek the solution $\mathbf{y} = A^{-1}\mathbf{b}$ of a nonsingular linear system $A\mathbf{y} = \mathbf{b}$ of n equations where the real matrix A has a small positive numerical nullity r. Then randomized additive preprocessing $A \Longrightarrow C = A + UV^T$ for $U, V \in \mathcal{G}_{0,\sigma}^{n\times r}$ for $\sigma \approx ||A||$ is expected to produce a well conditioned matrix C. We can strengthen this expectation by applying derandomization of Section 5.3.

The flowchart below combines such a preprocessing with the SMW formula and iterative refinement to compute the solution vector $\mathbf{y}$. Before specifying it we note that the solution of the original linear system $A\mathbf{y} = \mathbf{b}$ is reduced to accurate solution of $r+1$ linear systems of equations with matrix C. Indeed post-multiply the SMW formula by a vector $\mathbf{b}$ and obtain that

$$A^{-1}\mathbf{b} = C^{-1}\mathbf{b} + C^{-1}UG^{-1}V^HC^{-1}\mathbf{b} \quad \text{for} \quad G = I_r - V^HC^{-1}U.$$

Substitute $W_U = C^{-1}U$ and $\mathbf{w_b} = C^{-1}\mathbf{b}$ and obtain $A^{-1}\mathbf{b} = \mathbf{w_b} + W_UG^{-1}V^H\mathbf{w_b}$ for $G = I_r - V^HW_U$. This reduces the solution of a linear system $A\mathbf{y} = \mathbf{b}$ essentially to the solution of the matrix equation $CW_U = U$ and the linear system $C\mathbf{w_b} = \mathbf{b}$, computing the above matrix G, and its inversion. We can combine the equations $CW_U = U$ and $C\mathbf{w_b} = \mathbf{b}$ into the single matrix equation

$$CW = (U \mid \mathbf{b}) \text{ for } W = (W_U \mid W_\mathbf{b}). \tag{5.2}$$

Flowchart 5.1. Randomized Solution of a Linear System with Iterative Refinement

INPUT: *a vector $\mathbf{b}$ of a dimension n and an $n\times n$ ill conditioned matrix A having a small positive numerical nullity r.*

OUTPUT: *an approximate solution $\tilde{\mathbf{y}} \approx \mathbf{y}$ of the linear system $A\mathbf{y} = \mathbf{b}$.*

COMPUTATIONS*:*

1. *Apply additive preprocessing $A \Longrightarrow C = A + UV^T$ for a pair of scaled $n \times r$ random Gaussian matrices U and V. (With a high probability the matrix C is expected to be nonsingular and well conditioned).*

2. *Apply Gaussian elimination (or another direct algorithm) involving $O(n^3)$ flops to compute an approximate inverse $X \approx C^{-1}$. (Perform the computations by using the IEEE standard single or double precision. Application of the same algorithm to the original ill conditioned linear system $A\mathbf{y} = \mathbf{b}$ would require about as many flops but in extended precision).*

3. *Employ this inverse as the basis for iterative refinement to compute sufficiently accurate solutions of $r + 1$ auxiliary linear systems of equations with the matrix C and then recover the vector $\mathbf{y} = A^{-1}\mathbf{b}$ via the SMW formula.*

Elaboration upon and the analysis of this flowchart are quite involved and require extended precision for computing the Schur complement $G = I_r - V^HC^{-1}U$. This is done in a separate paper [Pa], but next we will briefly comment on interesting impact on the complexity of ill conditioned linear systems of equations.

Every loop of iterative refinement produces order $p - \log_2 \kappa(C)$ new correct bits per output value and is essentially reduced to multiplication of the matrices C and X by two vectors, that is to $4n^2 - 2n$ flops, which can be performed in a low precision p. Overall our randomized Flowchart 5.1 is expected to run by a factor $\frac{n}{r\log r}$ faster than the customary algorithms such as Gaussian elimination. In the case of a small positive integer r and a large condition number $\kappa(A)$ the flowchart uses nearly optimal number of bitwise operations, within polylogarithmic factors from an information lower bound.

Instead of iterative refinement one can apply other iterative algorithms such as CG or GMRES algorithms that involve no approximate inverse. This is importnat, e.g., in the case of multilevel Toeplitz or Hankel matrices, which can be multiplied by vectors fast but are inverted slowly. The CG or GMRES algorithms, however, are more sensitive to the success of preconditioning. In particular every CG iteration loop essentially amounts to multiplication of the matrices C and C^T by two vectors and produces order of $1/\kappa(C)$ new correct bits per an output value. Stronger bounds on $\kappa(C)$ are required to preserve this progress in the presence of rounding errors.

Iterative refinement and CG algorithms compute $O(rn)$ auxiliary values with high accuracy by accumulating them as the sums of sufficiently many low precision summands, in a way quite customary in symbolic lifting [P09/11].

5.8 Randomized structured preprocessing

Would the preprocessed matrices $C = A + UV^H$ and $K = \begin{pmatrix} A & -U \\ WV^H & W \end{pmatrix}$ inherit the structure of a matrix A having a small numerical nullity r? For a small value r the adverse impact of involving the $O(nr)$ entries of the matrices U, V and W on the structure is small, but it could be even smaller if these matrices have structure consistent with the matrix A. In this case we do not have formal extension of our basic Theorem 3.1 as well as all its corollaries, but we have consistent empirical support for such an extension in the case of Gaussian random Toeplitz matrices U, V and W and actually even where we imposed additional patterns of sparseness under very weak randomization conditions (see [PIMR10]).

Likewise we observed consistent preconditioning power stated in Corollary 3.4 for randomized multipliers even when we applied circulant multipliers filling their fisrt columns with ones and -1 and limiting randomization to the choice of the signs $+$ or $-$.

6 A randomized Toeplitz solver

The following theorem expresses the inverse of a nonsingular Toeplitz matrix $T = (t_{i-j})_{i,j=1}^{n}$ via two columns $K^{-1}\mathbf{e}_1$ and $K^{-1}\mathbf{e}_{n+1}$ of the inverse K^{-1} of an $(n+1) \times (n+1)$ Toeplitz matrix K that embeds T as its block submatrix.

Theorem 6.1. *Suppose $K = (t_{i,j})_{i,j=0}^{n}$ is a nonsingular $(n+1) \times (n+1)$ Toeplitz matrix, write $T = (t_{i,j})_{i,j=0}^{n-1}$, $\widehat{\mathbf{v}} = (v_i)_{i=0}^{n} = K^{-1}\mathbf{e}_1$, $\mathbf{v} = (v_i)_{i=0}^{n-1}$, $\mathbf{v}' = (v_i)_{i=1}^{n}$, $\widehat{\mathbf{w}} = (w_i)_{i=0}^{n} = K^{-1}\mathbf{e}_{n+1}$, $\mathbf{w} = (w_i)_{i=0}^{n-1}$, and $\mathbf{w}' = (w_i)_{i=1}^{n}$. (a) If $v_0 \neq 0$, then the matrix $T = (t_{i,j})_{i,j=0}^{n-1}$ is nonsingular and $v_0 T^{-1} = Z(\mathbf{v})Z^T(J\mathbf{w}') - Z(\mathbf{w})Z^T(J\mathbf{v}')$. (b) If $v_n \neq 0$, then the matrix $T_{10} = (t_{i,j})_{i=1,j=0}^{n,n-1}$ is nonsingular and $v_n T^{-1} = Z(\mathbf{w})Z^T(J\mathbf{v}') - Z(\mathbf{v})Z^T(J\mathbf{w}')$.*

Proof. Part (a) was proved in [GS72], part (b) in [GK72]; [BGY80, Theorem 7]) reproduces part (b). □

The theorem extends the better known formulae in [GS72], [H79], [HR84], [T90], each expressing T^{-1} via the solution of a pair of linear systems of equations with the matrix T itself.

In the case of a nonsingular real symmetric matrix K the first and the last columns of the matrix K^{-1} turn into one another up to reflection, that is $K^{-1}\mathbf{e}_1 = J_{n+1}K^{-1}\mathbf{e}_{n+1}$, because in this case the inverse K^{-1} is both symmetric and persymmetric. Then part (a) of Theorem 6.1 expresses the matrix T^{-1} via the first column of the matrix K^{-1} alone.

Let us apply Theorem 6.1 to support our randomized augmentation techniques for solving a nonsingular Toeplitz linear system $T\mathbf{y} = \mathbf{b}$ of n equations provided the matrix T has numerical nullity one.

To compute the solution vector $\mathbf{y} = T^{-1}\mathbf{b}$, we first embed the matrix T into an $(n+1) \times (n+1)$ Toeplitz matrix $K = \begin{pmatrix} w & \mathbf{v}^T \\ \mathbf{f} & T \end{pmatrix}$. By virtue of the Toeplitz structure of the matrix K we have $w = \mathbf{e}_1^T T \mathbf{e}_1$, and the vectors $\mathbf{f} = (f_i)_{i=1}^{n}$ and $\mathbf{v} = (v_i)_{i=1}^{n}$ are filled with the respective entries of the

matrix T except for the two coordinates f_n and v_n, which we choose at random and then scale to have the ratio $\frac{||K||}{||T||}$ neither large nor small (cf. [GS72]).

By virtue of Corollary 2.4 this policy is likely to produce a nonsingular matrix K whose inverse is likely to have a nonzero entry $\mathbf{e}_1^T K^{-1} \mathbf{e}_1$. These two implications of Corollary 2.4 were in good accordance with our test results, in which the matrix K was also consistently well conditioned, even though we used only two random parameters.

Part (a) of Theorem 6.1 expresses the inverse T^{-1} via the first column $\mathbf{v} = K^{-1}\mathbf{e}_1$ and the last column $\mathbf{w} = K^{-1}\mathbf{e}_{n+1}$ of the inverse matrix K^{-1}.

To summarize, we reduce the solution of the original ill conditioned Toeplitz linear system $T\mathbf{y} = \mathbf{b}$ to computing highly accurate solutions of two linear systems $K\mathbf{x} = \mathbf{e}_1$ and $K\mathbf{z} = \mathbf{e}_{n+1}$ both expected to be well conditioned. High accuracy is needed to counter magnification of the input and rounding errors, expected in the case of ill conditioned input.

To solve the two latter systems, we first employ the effective algorithms in [KV99], [V99], [VBHK01], and [VK98] and then apply iterative refinement with double precision. We refer to the resulting algorithm as **Algorithm 6.1**.

One can readily extend the approach of this section to the case of Toeplitz-like, Hankel and Hankel-like inputs.

In the important special case where the Toeplitz matrix T is real symmetric, we can choose real scalar w and a single real vector $\mathbf{f} = \mathbf{v}$ to yield a real symmetric matrix $K = \begin{pmatrix} w & \mathbf{v}^T \\ \mathbf{v} & T \end{pmatrix}$. Then Algorithm 6.1 is simplified because $\mathbf{w} = K^{-1}\mathbf{e}_{n+1} = J_{n+1}\mathbf{v} = J_{n+1}K^{-1}\mathbf{e}_1$, and we only need to solve a single linear system with the matrix K. In Section 7.6 we test the resulting algorithm for solving an ill conditioned real symmetric Toeplitz linear system.

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 was compiled with the GNU gfortran compiler within the Cygwin environment. Random numbers were generated with the random_number intrinsic Fortran function, assuming the uniform probability distribution over the range $\{x : -1 \le x < 1\}$.

7.1 Conditioning tests

We computed the condition numbers of $n \times n$ random general matrices for $n = 2^k$, $k = 5, 6, \dots$, with the entries sampled in the range $[-1, 1)$ as well as complex general, Toeplitz, and circulant matrices whose entries had real and imaginary parts sampled at random in the same range $[-1, 1)$. We performed 100 tests for each dimension n and represented the test results in Tables 7.2–7.4. The last four columns of each table display the average (mean), minimum, maximum, and standard deviation of the computed condition numbers of the input matrices, respectively. Specifically we computed the values $\kappa(A) = ||A||\,||A^{-1}||$ for general and circulant matrices A and the values $\kappa_1(A) = ||A||_1\,||A^{-1}||_1$ for Toeplitz matrices A. We computed and displayed in Table 7.3 the 1-norms of Toeplitz matrices and their inverses rather than their 2-norms to facilitate the computations in the case of inputs of large sizes. Table 7.1 shows that the 1-norms and 2-norms are quite close to each other. It displays the data on $n \times n$ general, Toeplitz, and circulant matrices A for $n = 32, 64, \dots, 1024$. We sampled the matrix entries at random in the range of $-1 \le x < 1$ and performed 100 conditioning tests for each matrix class and each size.

7.2 Preconditioning tests

Table 7.5 reproduces some results of testing the preconditioning power of additive preprocessing in [PIMR10]. We tested input matrices of the following classes.

1n. *Nonsymmetric matrices of type I with numerical nullity ν.* $A = S\Sigma_\nu T^H$ are $n \times n$ matrices where G and H are $n \times n$ random orthogonal matrices, that is, the factors Q in the QR factorizations of random real matrices; $\Sigma_\nu = \mathrm{diag}(\sigma_j)_{j=1}^n$ is the diagonal matrix such that $\sigma_{j+1} \le \sigma_j$ for $j = 1, \dots, n-1$, $\sigma_1 = 1$, the values $\sigma_2, \dots, \sigma_{n-\nu-1}$ are randomly sampled in the semi-open interval $[0.1, 1)$, $\sigma_{n-\nu} = 0.1$, $\sigma_j = 10^{-16}$ for $j = n-\nu+1, \dots, n$, and therefore $\kappa(A) = 10^{16}$ [H02, Section 28.3].

1s. *Symmetric matrices of type I with numerical nullity ν.* The same as in part 1n, but for $G = H$.

The matrices of six other classes were constructed in the form of $\frac{A}{||A||} + \beta I$ where the recipes for defining the matrices A and scalars β are specified below.

2n. *Nonsymmetric matrices of type II with numerical nullity ν.* $A = (W \mid WZ)$ where W and Z are random orthogonal matrices of sizes $n \times (n-\nu)$ and $(n-\nu) \times \nu$, respectively.

2s. *Symmetric matrices of type II with numerical nullity ν.* $A = WW^H$ where W are random orthogonal matrices of size $n \times (n-\nu)$.

3n. *Nonsymmetric Toeplitz-like matrices with numerical nullity ν.* $A = c(T \mid TS)$ for random Toeplitz matrices T of size $n \times (n-\nu)$ and S of size $(n-\nu) \times \nu$ and for a positive scalar c such that $||A|| \approx 1$.

3s. *Symmetric Toeplitz-like matrices with numerical nullity ν.* $A = cTT^H$ for random Toeplitz matrices T of size $n \times (n-\nu)$ and a positive scalar c such that $||A|| \approx 1$.

4n. *Nonsymmetric Toeplitz matrices with numerical nullity one.* $A = (a_{i,j})_{i,j=1}^n$ is an $n \times n$ Toeplitz matrix. Its entries $a_{i,j} = a_{i-j}$ are random for $i - j < n - 1$. The entry $a_{n,1}$ is selected to ensure that the last row is linearly expressed through the other rows.

4s. *Symmetric Toeplitz matrices with numerical nullity one.* $A = (a_{i,j})_{i,j=1}^n$ is an $n \times n$ Toeplitz matrix. Its entries $a_{i,j} = a_{i-j}$ are random for $|i - j| < n - 1$, whereas the entry $a_{1,n} = a_{n,1}$ is a root of the quadratic equation $\det A = 0$. We have repeatedly generated the matrices A until we arrived at the quadratic equation having real roots.

The scalar β was set equal to 10^{-16} for the symmetric matrices A, in the classes 2s, 3n, and 4s, so that $\kappa(A) = 10^{16} + 1$ in these cases. For the nonsymmetric matrices A the scalar β was defined by an iterative process such that $||A|| \approx 1$ and $10^{-18}||A|| \le \kappa(A) \le 10^{-16}||A||$ [PIMR10, Section 8.2].

Table 7.5 displays the average values of the condition numbers $\kappa(C)$ for the matrices $C = A + UU^T$ over 100,000 tests for the inputs in the above classes, $\nu = r$ in the range $\{1, 2, 4, 8\}$ and $n = 100$. The additive preprocessor UU^T was defined by a normalized $n \times r$ matrix $U = U/||U||$ where $U^T = (\pm I \mid O_{r,r} \mid \pm I \mid O_{r,r} \mid \dots \mid O_{r,r} \mid \pm I \mid O_{r,s})$, the integer s was chosen to obtain $n \times r$ matrices U, and the signs for the matrices $\pm I$ were chosen at random.

In our further tests the condition numbers of the matrices $C = A + 10^p UV^T$ for $p = -10, -5, 5, 10$ were steadily growing within a factor $10^{|p|}$ as the value $|p|$ was growing. This showed the importance of proper scaling of the additive preprocessor UV^T.

7.3 Solution of general linear systems of equations with random circulant multipliers

Table 7.6 (cf. [PQZa, Table 2]) displays the results of our tests of the solution of well conditioned linear systems $A\mathbf{y} = \mathbf{b}$ of n equations whose coefficient matrix had an ill conditioned $n/2 \times n/2$ submatrix for n ranging from 64 to 1024. We 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 output corrupted solutions with the residual norms ranging from 10 to 10^8. When we preprocessed the systems with circulant multipliers filled with ones and -1 (with the signs $\pm$ chosen at random), the norms decreased to at worst 10^{-7} for all inputs. Table 7.6 also shows further decrease of the norm in a single step of iterative refinement.

7.4 Approximation of the tails of SVDs and lower-rank approximation of a matrix

Table 7.7 (see [PQ10, Section 10.6]) displays the data from our tests on approximation of trailing singular spaces of the SVD of an $n \times n$ matrix A having numerical nullity r and on approximation of this matrix with a matrix of rank $n - r$.

For $n = 64, 128, 256$ we generated pairs of $n \times n$ random unitary matrices S and T and diagonal matrices $\Sigma = \mathrm{diag}(\sigma_j)_{j=1}^n$ such that $\sigma_j = 1/j,\ j = 1, \dots, n-r$, $\sigma_j = 10^{-10},\ j = n-r+1, \dots, n$. Then we computed the input matrices $A = S\Sigma T^T$ (with $\mathrm{cond}\, A = 10^{10}$) as well as the matrix bases $T_r = T \begin{pmatrix} 0 \\ I_r \end{pmatrix}$ for the trailing singular spaces of these matrices. We also generated pairs of $n \times r$ random matrices U and V for $r = 1, 8, 32$, then scaled them to have the ratios $||UV^H||/||A||$ neither large nor small, and computed the matrices $C = A + UV^T$, $B_r = C^{-1}U$, AB_r, Y, B_rY, $B_rY - T_r$, $Q = Q(B_r)$, and $AQQ^H = A - A(I_n - QQ^H)$ where the matrices Y minimized the norms $||B_rY - T_r||$.

Table 7.7 displays the data on the values $\mathrm{cond}\, A$ and the relative residual norms $\mathrm{rrn}_1 = \frac{||B_rY - T_r||}{||B_rY||}$, $\mathrm{rrn}_2 = \frac{||AB_r||}{||A||\ ||B_r||}$, and $\mathrm{rrn}_3 = \frac{||AQQ^H||}{||A||}$ obtained in 100 runs of our tests.

7.5 Solution of general linear systems of equations via approximation of trailing singular spaces of the SVDs

We chose $n = 32, 64$ and $r = 1, 2, 4$ and for every pair (n, r) generated 100 instances of vectors $\mathbf{b}$ and matrices A, U, and V as follows.

We generated (a) random vectors $\mathbf{b}$ of dimension n, (b) the matrices A as the error-free products $S\Sigma T^H$ where S and T were $n \times n$ random real orthonormal matrices (generated with double precision), $\Sigma = \mathrm{diag}(\sigma_j)_{j=1}^n$, $\sigma_{n-j} = 10^{j-17}$ for $j = 1, \dots, r$, and $\sigma_{n-j} = 1/(n-j)$ for $j = r+1, \dots, n-1$ [H02, Section 28.3], and (c) $n \times r$ random matrices U and V such that $||U| = ||A||$ and $||V|| = 1$.

For every choice of these matrices we solved the linear systems $A\mathbf{y} = \mathbf{b}$ based on Algorithm 5.1. We first generated $n \times (n-r)$ random matrices K_0 and L_0 and then computed the matrices $C = A + UV^T$ (which were always nonsingular and well conditioned in our tests), $K_1 = C^{-T}V$, $L_1 = C^{-1}U$, and $W = (K_0 \mid K_1)^T A (L_0 \mid L_1) = \begin{pmatrix} W_{00} & W_{01} \\ W_{10} & W_{11} \end{pmatrix}$. In all our tests the $(n-r) \times (n-r)$ leading principal $(n-r) \times (n-r)$ block $W_{00} = K_0^T A L_0$ was strongly well conditioned and strongly dominated the three other blocks W_{01}, W_{10}, and W_{11} in the 2×2 block matrix W, as we expected to see based on our analysis in Section 5.4. We computed the dominated blocks W_{01}, W_{10}, and W_{11} with extended precision. Then we solved the linear system $W\mathbf{x} = (K_0 \mid K_1)^T \mathbf{b}$. We first applied Gaussian elimination with no pivoting to eliminate the subdiagonal block. Then we readily computed the solution of the resulting block triangular linear system, whose both diagonal blocks were expected and indeed consistently turned out to be much better conditioned than the original matrix A. Finally we computed and output the vector $\mathbf{y} = (L_0 \mid L_1)\mathbf{x}$.

Table 7.8 shows the average (mean), minimum and maximum values of the relative residual norms $||A\mathbf{y} - \mathbf{b}||/||\mathbf{b}||$ of the output vectors $\mathbf{y}$ as well as the standard deviations observed in these tests.

For the same ill conditioned inputs the Subroutine MLDIVIDE(A,B) for Gaussian elimination from MATLAB produced corrupted outputs, as can be seen from in Table 7.9.

7.6 Solution of a real symmetric Toeplitz linear system of equations with randomized augmentation

We solved 100 real symmetric linear systems of equations $T\mathbf{y} = \mathbf{b}$ for each n where we used vectors $\mathbf{b}$ with random coordinates from the range $[-1, 1)$ and Toeplitz matrices $T = S + 10^{-9} I_n$ for an $n \times n$ singular symmetric Toeplitz matrices S with nullity one, generated according to the recipe in [PQ10, Section 10.1b].

Table 7.10 shows the average CPU time of the solution by our Algorithm 6.1 and, for comparison, based on the QR factorization and SVD, which we computed by applying the LAPACK procedures DGEQRF and DGESVD, respectively.

The abbreviations "Alg. 6.1", "QR", and "SVD" indicate the respective algorithms. The last two columns of the table display the ratios of these data on the CPU time.

We measured the CPU time with the mclock function by counting cycles. One can convert them into seconds by dividing their number by a constant CLOCKS_PER_SEC, which is 1000 on our platform. The table entries are marked by a "-" where the tests required too long runtime and were not completed.

We obtained the solutions $\mathbf{y}$ with the relative residual norms of about 10^{-15} in all three algorithms, which showed that Algorithm 6.1 employing iterative refinement was as reliable as the QR and SVD based solutions but ran much faster.

We refer the reader to [PQZC, Table 3] on similar test results on the solution of ill conditioned homogeneous Toeplitz linear systems.

Table 7.1: Norms of random general, Toeplitz and circulant matrices and of their inverses

matrix A	n	$\|\|A\|\|_1$	$\|\|A\|\|_2$	$\frac{\|\|A\|\|_1}{\|\|A\|\|_2}$	$\|\|A^{-1}\|\|_1$	$\|\|A^{-1}\|\|_2$	$\frac{\|\|A^{-1}\|\|_1}{\|\|A^{-1}\|\|_2}$
General	32	1.9×10^1	1.8×10^1	1.0×10^0	4.0×10^2	2.1×10^2	1.9×10^0
General	64	3.7×10^1	3.7×10^1	1.0×10^0	1.2×10^2	6.2×10^1	2.0×10^0
General	128	7.2×10^1	7.4×10^1	9.8×10^{-1}	3.7×10^2	1.8×10^2	2.1×10^0
General	256	1.4×10^2	1.5×10^2	9.5×10^{-1}	5.4×10^2	2.5×10^2	2.2×10^0
General	512	2.8×10^2	3.0×10^2	9.3×10^{-1}	1.0×10^3	4.1×10^2	2.5×10^0
General	1024	5.4×10^2	5.9×10^2	9.2×10^{-1}	1.1×10^3	4.0×10^2	2.7×10^0
Toeplitz	32	1.8×10^1	1.9×10^1	9.5×10^{-1}	2.2×10^1	1.3×10^1	1.7×10^0
Toeplitz	64	3.4×10^1	3.7×10^1	9.3×10^{-1}	4.6×10^1	2.4×10^1	2.0×10^0
Toeplitz	128	6.8×10^1	7.4×10^1	9.1×10^{-1}	1.0×10^2	4.6×10^1	2.2×10^0
Toeplitz	256	1.3×10^2	1.5×10^2	9.0×10^{-1}	5.7×10^2	2.5×10^2	2.3×10^0
Toeplitz	512	2.6×10^2	3.0×10^2	8.9×10^{-1}	6.9×10^2	2.6×10^2	2.6×10^0
Toeplitz	1024	5.2×10^2	5.9×10^2	8.8×10^{-1}	3.4×10^2	1.4×10^2	2.4×10^0
Circulant	32	1.6×10^1	1.8×10^1	8.7×10^{-1}	9.3×10^0	1.0×10^1	9.2×10^{-1}
Circulant	64	3.2×10^1	3.7×10^1	8.7×10^{-1}	5.8×10^0	6.8×10^0	8.6×10^{-1}
Circulant	128	6.4×10^1	7.4×10^1	8.6×10^{-1}	4.9×10^0	5.7×10^0	8.5×10^{-1}
Circulant	256	1.3×10^2	1.5×10^2	8.7×10^{-1}	4.7×10^0	5.6×10^0	8.4×10^{-1}
Circulant	512	2.6×10^2	3.0×10^2	8.7×10^{-1}	4.5×10^0	5.4×10^0	8.3×10^{-1}
Circulant	1024	5.1×10^2	5.9×10^2	8.7×10^{-1}	5.5×10^0	6.6×10^0	8.3×10^{-1}

Table 7.2: condition numbers $\kappa(A)$ of random matrices A

n	**input**	**min**	**max**	**mean**	**std**
32	real	2.4×10^1	1.8×10^3	2.4×10^2	3.3×10^2
32	complex	2.7×10^1	8.7×10^2	1.1×10^2	1.1×10^2
64	real	4.6×10^1	1.1×10^4	5.0×10^2	1.1×10^3
64	complex	5.2×10^1	4.2×10^3	2.7×10^2	4.6×10^2
128	real	1.0×10^2	2.7×10^4	1.1×10^3	3.0×10^3
128	complex	1.3×10^2	2.5×10^3	3.9×10^2	3.3×10^2
256	real	2.4×10^2	8.4×10^4	3.7×10^3	9.7×10^3
256	complex	2.5×10^2	1.4×10^4	1.0×10^3	1.5×10^3
512	real	3.9×10^2	7.4×10^5	1.8×10^4	8.5×10^4
512	complex	5.7×10^2	3.2×10^4	2.3×10^3	3.5×10^3
1024	real	8.8×10^2	2.3×10^5	8.8×10^3	2.4×10^4
1024	complex	7.2×10^2	1.3×10^5	5.4×10^3	1.4×10^4
2048	real	2.1×10^3	2.0×10^5	1.8×10^4	3.2×10^4
2048	complex	2.3×10^3	5.7×10^4	6.7×10^3	7.2×10^3

Table 7.3: condition numbers $\kappa_1(A)$ of random Toeplitz matrices A

n	**min**	**mean**	**max**	**std**
256	9.1×10^2	9.2×10^3	1.3×10^5	1.8×10^4
512	2.3×10^3	3.0×10^4	2.4×10^5	4.9×10^4
1024	5.6×10^3	7.0×10^4	1.8×10^6	2.0×10^5
2048	1.7×10^4	1.8×10^5	4.2×10^6	5.4×10^5
4096	4.3×10^4	2.7×10^5	1.9×10^6	3.4×10^5
8192	8.8×10^4	1.2×10^6	1.3×10^7	2.2×10^6

Table 7.4: condition numbers $\kappa(A)$ of random circulant matrices A

n	**min**	**mean**	**max**	**std**
256	9.6×10^0	1.1×10^2	3.5×10^3	4.0×10^2
512	1.4×10^1	8.5×10^1	1.1×10^3	1.3×10^2
1024	1.9×10^1	1.0×10^2	5.9×10^2	8.6×10^1
2048	4.2×10^1	1.4×10^2	5.7×10^2	1.0×10^2
4096	6.0×10^1	2.6×10^2	3.5×10^3	4.2×10^2
8192	9.5×10^1	3.0×10^2	1.5×10^3	2.5×10^2
16384	1.2×10^2	4.2×10^2	3.6×10^3	4.5×10^2
32768	2.3×10^2	7.5×10^2	5.6×10^3	7.1×10^2
65536	2.4×10^2	1.0×10^3	1.2×10^4	1.3×10^3
131072	3.9×10^2	1.4×10^3	5.5×10^3	9.0×10^2
262144	6.3×10^2	3.7×10^3	1.1×10^5	1.1×10^4
524288	8.0×10^2	3.2×10^3	3.1×10^4	3.7×10^3
1048576	1.2×10^3	4.8×10^3	3.1×10^4	5.1×10^3

Table 7.5: Preconditioning tests

Type	$\nu = r$	Cond (C)
1n	1	3.21E+2
1n	2	4.52E+3
1n	4	2.09E+5
1n	8	6.40E+2
1s	1	5.86E+2
1s	2	1.06E+4
1s	4	1.72E+3
1s	8	5.60E+3
2n	1	8.05E+1
2n	2	6.82E+3
2n	4	2.78E+4
2n	8	3.59E+3
2s	1	1.19E+3
2s	2	1.96E+3
2s	4	1.09E+4
2s	8	9.71E+3
3n	1	2.02E+4
3n	2	1.53E+3
3n	4	6.06E+2
3n	8	5.67E+2
3s	1	2.39E+4
3s	2	2.38E+3
3s	4	1.69E+3
3s	8	6.74E+3
4n	1	4.93E+2
4n	2	4.48E+2
4n	4	2.65E+2
4n	8	1.64E+2
4s	1	1.45E+3
4s	2	5.11E+2
4s	4	7.21E+2
4s	8	2.99E+2

Table 7.6: relative residual norms of the solutions by GENP with randomized cirulant multiplicative preprocessing

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 7.7: approximation of tails of the SVDs and lower-rank approximation of a matrix (cf. [PQ10])

r	cond(A) or rrn_i	n	**min**	**max**	**mean**	**std**
1	cond(A)	64	$2.38 \times 10^{+02}$	$1.10 \times 10^{+05}$	$6.25 \times 10^{+03}$	$1.68 \times 10^{+04}$
1	cond(A)	128	$8.61 \times 10^{+02}$	$7.48 \times 10^{+06}$	$1.32 \times 10^{+05}$	$7.98 \times 10^{+05}$
1	cond(A)	256	$9.70 \times 10^{+02}$	$3.21 \times 10^{+07}$	$3.58 \times 10^{+05}$	$3.21 \times 10^{+06}$
1	rrn_1	64	4.01×10^{-10}	1.50×10^{-07}	5.30×10^{-09}	1.59×10^{-08}
1	rrn_1	128	7.71×10^{-10}	5.73×10^{-07}	1.58×10^{-08}	6.18×10^{-08}
1	rrn_1	256	7.57×10^{-10}	3.2×10^{-07}	1.69×10^{-08}	5.02×10^{-08}
1	rrn_2	64	1.07×10^{-08}	4.71×10^{-06}	1.46×10^{-07}	4.90×10^{-07}
1	rrn_2	128	3.64×10^{-08}	3.05×10^{-05}	8.35×10^{-06}	3.29×10^{-06}
1	rrn_2	256	8.25×10^{-08}	3.30×10^{-05}	1.72×10^{-06}	5.03×10^{-06}
1	rrn_3	64	4.01×10^{-10}	1.50×10^{-07}	5.30×10^{-09}	1.59×10^{-08}
1	rrn_3	128	7.71×10^{-10}	5.73×10^{-07}	1.58×10^{-08}	6.18×10^{-08}
1	rrn_3	256	7.57×10^{-10}	3.22×10^{-07}	1.69×10^{-08}	5.02×10^{-08}
8	cond(A)	64	$1.26 \times 10^{+03}$	$1.61 \times 10^{+07}$	$2.68 \times 10^{+05}$	$1.71 \times 10^{+06}$
8	cond(A)	128	$2.92 \times 10^{+03}$	$3.42 \times 10^{+06}$	$1.58 \times 10^{+05}$	$4.12 \times 10^{+05}$
8	cond(A)	256	$1.39 \times 10^{+04}$	$8.75 \times 10^{+07}$	$1.12 \times 10^{+06}$	$8.74 \times 10^{+06}$
8	rrn_1	64	3.39×10^{-10}	2.27×10^{-06}	2.74×10^{-08}	2.27×10^{-07}
8	rrn_1	128	4.53×10^{-10}	1.91×10^{-07}	1.03×10^{-08}	2.79×10^{-08}
8	rrn_1	256	8.74×10^{-10}	1.73×10^{-07}	7.86×10^{-09}	1.90×10^{-08}
8	rrn_2	64	3.90×10^{-08}	1.47×10^{-04}	1.79×10^{-06}	1.47×10^{-05}
8	rrn_2	128	9.56×10^{-08}	2.97×10^{-05}	1.50×10^{-06}	4.12×10^{-06}
8	rrn_2	256	2.99×10^{-07}	3.91×10^{-05}	2.56×10^{-06}	5.70×10^{-06}
8	rrn_3	64	1.54×10^{-09}	7.59×10^{-06}	8.87×10^{-08}	7.58×10^{-07}
8	rrn_3	128	1.82×10^{-09}	7.27×10^{-07}	2.95×10^{-08}	8.57×10^{-08}
8	rrn_3	256	2.62×10^{-09}	3.89×10^{-07}	2.27×10^{-08}	5.01×10^{-08}
32	cond(A)	64	$1.77 \times 10^{+03}$	$9.68 \times 10^{+06}$	$1.58 \times 10^{+05}$	$9.70 \times 10^{+05}$
32	cond(A)	128	$1.65 \times 10^{+04}$	$6.12 \times 10^{+07}$	$1.02 \times 10^{+06}$	$6.19 \times 10^{+06}$
32	cond(A)	256	$3.57 \times 10^{+04}$	$2.98 \times 10^{+08}$	$4.12 \times 10^{+06}$	$2.98 \times 10^{+07}$
32	rrn_1	64	2.73×10^{-10}	3.29×10^{-08}	2.95×10^{-09}	4.93×10^{-09}
32	rrn_1	128	3.94×10^{-10}	1.29×10^{-07}	7.18×10^{-09}	1.64×10^{-08}
32	rrn_1	256	6.80×10^{-10}	4.00×10^{-07}	1.16×10^{-08}	4.27×10^{-08}
32	rrn_2	64	2.59×10^{-08}	2.11×10^{-06}	2.07×10^{-07}	3.29×10^{-07}
32	rrn_2	128	1.45×10^{-07}	1.82×10^{-05}	1.50×10^{-06}	2.76×10^{-06}
32	rrn_2	256	3.84×10^{-07}	7.06×10^{-05}	5.27×10^{-06}	1.14×10^{-05}
32	rrn_3	64	2.10×10^{-09}	1.49×10^{-07}	1.55×10^{-08}	2.18×10^{-08}
32	rrn_3	128	2.79×10^{-09}	3.80×10^{-07}	3.81×10^{-08}	6.57×10^{-08}
32	rrn_3	256	5.35×10^{-09}	1.05×10^{-06}	5.70×10^{-08}	1.35×10^{-07}

Table 7.8: Relative residual norms for a linear system of equations via nmb approximation

n	**r**	**min**	**max**	**mean**	**std**
32	1	1.49×10^{-13}	1.36×10^{-9}	4.25×10^{-11}	1.56×10^{-10}
32	2	3.70×10^{-13}	2.13×10^{-8}	3.83×10^{-10}	2.35×10^{-9}
32	4	9.33×10^{-13}	1.08×10^{-8}	3.37×10^{-10}	1.26×10^{-9}
64	1	1.11×10^{-12}	6.87×10^{-9}	2.03×10^{-10}	7.49×10^{-10}
64	2	1.53×10^{-12}	1.21×10^{-8}	5.86×10^{-10}	1.77×10^{-9}
64	4	2.21×10^{-12}	1.27×10^{-7}	1.69×10^{-9}	1.28×10^{-8}

Table 7.9: Relative residual norms for a linear system of equations with MLDIVIDE(A,B)

n	**r**	**min**	**max**	**mean**	**std**
32	1	6.34×10^{-3}	7.44×10^{1}	1.74×10^{0}	7.53×10^{0}
32	2	2.03×10^{-2}	1.32×10^{1}	9.19×10^{-1}	1.62×10^{0}
32	4	4.57×10^{-2}	1.36×10^{1}	1.14×10^{0}	1.93×10^{0}
64	1	3.82×10^{-3}	9.93×10^{0}	1.03×10^{0}	1.66×10^{0}
64	2	1.96×10^{-2}	1.27×10^{2}	3.09×10^{0}	1.40×10^{1}
64	4	7.13×10^{-3}	6.63×10^{0}	8.23×10^{-1}	1.20×10^{0}

Table 7.10: CPU time (in cycles) for solving an ill conditioned real symmetric Toeplitz linear system

n	**Alg. 6.1**	**QR**	**SVD**	**QR/Alg. 6.1**	**SVD/Alg. 6.1**
512	56.3	148.4	4134.8	2.6	73.5
1024	120.6	1533.5	70293.1	12.7	582.7
2048	265.0	11728.1	–	44.3	–
4096	589.4	–	–	–	–
8192	1304.8	–	–	–	–

Appendix

A Newton's structured iteration and preconditioning

Recall Newton's iteration for matrix inversion

$$X_{i+1} = X_i(2I - CX_i),\ i = 0, 1, \dots. \tag{A.1}$$

Its ith loop squares the residual $I - CX_i$, that is, we have

$$I - CX_{i+1} = (I - CX_i)^2 = (I - CX_0)^{2^{i+1}}. \tag{A.2}$$

Therefore

$$||I - CX_{i+1}|| \le ||I - CX_i||^2 = ||I - CX_0||^{2^{i+1}},\ i = 0, 1, \dots, \tag{A.3}$$

so that the approximations X_i quadratically converge to the inverse C^{-1} right from the start provided that $||I - CX_0|| < 1$.

We can ensure that $||I - CX_0|| \le 1 - \frac{2n}{(\kappa(C))^2(1+n)}$ by choosing $X_0 = \frac{2nC^H}{(1+n)||C||_1||C||_\infty}$ [PS91].

Such a map $C \Longrightarrow X_0$ preserves the matrix structure of Toeplitz or Hankel type, but is the structure maintained throughout the iteration? Not automatically. In fact a Newton's loop can triple the displacement rank of the matrix X_k. The structure can be maintained, however, via recursive compression of the displacement (also called recompression), in which case we arrive at *Newton's structured iteration*. In particular we can periodically set to zero the smallest singular values of the displacements of the matrices X_i to keep the length of the displacements within a fixed tolerance t, equal to or a little exceeding the displacement rank of the input matrix C.

We refer the reader to [P01, Chapter 6] on the history, variations, and analysis of this approach, proposed in [P92], [P93], and [P93a]. According to the estimates in [P01], the structured iteration converges quadratically right from the start provided $||I - CX_0|| < \frac{1}{(1+||Z_e||+||Z_f||)\kappa(C)}||L^{-1}||$, $||L^{-1}|| \le c_{e,f}n$, L denotes the operator $\nabla_{Z_e,Z_f}(C)$ for $e \ne f$ or $\Delta_{Z_e,Z_f^T}(C)$ for $ef \ne 1$, and $c_{e,f}$ is a constant defined by e and f.

Newton's iteration can be incorporated into our randomized algorithms. E.g., it can substitute for Gussian elimination in Flowchart 5.1. One can also try to apply preconditioning to avoid or to accelerate the stage of slow start of Newton's iteration, observed where the initial residual norm $||I - CX_0||$ is close to one. Preconditioning is a natural way to decreasing this norm, that is, with our preconditioning we can ensure that $||I - CX_0|| \le u$ for a constant $u < 1$, then apply $O(\log n)$ loops of iterative refinement to satisfy the above initialization bound, and finally shift to Newton's structured iteration. We must perform extra refinement steps to yield the output with high accuracy, but these steps are noncostly in the case of structured Newton's iteration.

By applying iterative refinement at the initial stage of slow start, we put up with linear convergence, but the experiments reported in [P01, Table 6.21] suggest that we can avoid this stage in the case of Toeplitz matrices C. Namely these experiments show global convergence of Newton's structured iteration with compression (right from the start) in about 25% of tests, including the cases where the initial residual norm $||I - CX_0||$ was very close to one.

Motivated by these tests we propose concurrent applications of a number of variations of Newton's structured iteration, including variations of its compression policy in [PS91], [P01, Chapter 6], and [PVW04], to a number of scaled randomized small rank modifications and small size augmentations of the input matrix. As soon as one of these applications produces the inverse, we can readily recover the inverse of the original matrix via the SMW formula (also see Theorem 5.4 for augmetations).

Of course it is interesting whether this approach can also work for other classes of structured matrices.

References

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

[AHU74] A. V. Aho, J. E. Hopcroft, J. D. Ullman, *The Design and Analysis of Algorithms.* Addison-Wesley, Reading, MA, 1974.

[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.

[B85] D. Bini, Tensor and Border Rank of Certain Classes of Matrices and the Fast Evaluation of Determinant, Inverse Matrix and Eigenvalues, *Calcolo*, **22**, 209–228, 1985.

[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.

[BA80] R. R. Bitmead, B. D. O. Anderson, Asymptotically Fast Solution of Toeplitz and Related Systems of Linear Equations, *Linear Algebra and Its Applications*, **34**, 103–116, 1980.

[BC87] D. Bini, M. Capovani, Tensor Rank and Border Rank of Band Toeplitz Matrices, *SIAM J. on Computing*, **2**, 252–258, 1987.

[BCLR79] D. Bini, M. Capovani, G. Lotti, F. Romani $O(n^{2.7799})$ Complexity for $n \times n$ Approximate Matrix Multiplication, *Information Processing Letters*, **8**, 234–235, 1979.

[BEPP] H. Brönnimann, I. Z. Emiris, V. Y. Pan, S. Pion, Sign Determination in Residue Number Systems, *Theoretical Computer Science*, **210**, **1**, 173–197, 1999. Proceedings version in *Proceedings of 13th Annual ACM Symposium on Computational Geometry*, 174–182, ACM Press, New York, 1997.

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

[BGY80] R. P. Brent, F. G. Gustavson, D. Y. Y. Yun, Fast Solution of Toeplitz Systems of Equations and Computation of Padé Approximations, *J. Algorithms*, **1**, 259–295, 1980.

[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.

[BMP00] D. Bondyfalat, B. Mourrain, V. Y. Pan, Computation of a Specified Root of a Polynomial System of Equations Using Eigenvectors, *Linear Algebra and Its Applications*, **319**, 193–209 (2000). (Proceedings version in *ACM-SIGSAM ISSAC'98.*)

[BP94] D. Bini, V. Y. Pan, *Polynomial and Matrix Computations,* volume 1: Fundamental Algorithms, Birkhäuser, Boston, 1994.

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

[CDG03] B. Carpentieri, I.S. Duff, L. Giraud, A class of spectral two-level preconditioners, *SIAM J. Scientific Computing*, **25, 2**, 749–765, 2003.

[CGLX] S. Chandrasekaran, M. Gu, X. S. Li, J. Xia, Superfast Multifrontal Method for Large Structured Linear Systems of Equations, *SIAM. J. on Matrix Analysis and Applications*, **31**, 1382–1411, 2009.

[CGSXZ] S. Chandrasekaran, M. Gu, X. Sun, J. Xia, J. Zhu, A Superfast Algorithm for Toeplitz Systems of linear Equations, *SIAM. J. on Matrix Analysis and Applications*, **29**, **4**, 1247–1266, 2007.

[CPW74] R.E. Cline, R.J. Plemmons, and G. Worm, Generalized Inverses of Certain Toeplitz Matrices, *Linear Algebra and Its Applications,* **8**, 25–33, 1974.

[CW90] Coppersmith, S. Winograd, Matrix Multiplicaton via Arithmetic Progressions. *J. Symbolic Comput.*, **9, 3**, 251–280, 1990.

[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.

[DH03] J. Demmel, Y. Hida, Accurate and Efficient Floating Point Summation, *SIAM J. on Scientific Computing*, **25**, 1214–1248, 2003.

[DIS] C.-E. Drevet, M. N. Islam, E. Schost, Optimization Techniques for Small Matrix Multiplication, preprint, 2010.

[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.

[EG99] Y. Eidelman, I. Gohberg, On a New Class of Structured Matrices, *Integral Equations and Operator Theory*, **34**, 293–324, Birkhäuser, Basel, 1999.

[EP02] I. Z. Emiris, V. Y. Pan, Symbolic and Numerical Methods for Exploiting Structure in Constructing Resultant Matrices, *J. of Symbolic Computation*, **33**, 393-413, 2002. Proc. Version in *ISSAC 97*.

[EP03/05] I. Z. Emiris, V. Y. Pan, Improved Algorithms for Computing Determinants and Resultants, *J. of Complexity*, **21**, **1**, 43–71, 2005. Proceedings version in *Proceedings of 6th International Workshop on Computer Algebra in Scientific Computing (CASC '03)*, E. W. Mayr, V. G. Ganzha, E. V. Vorozhtzov (editors), 81–94, Technische Univ. München, Germany, 2003.

[EP10] I. Z. Emiris, V. Y. Pan, Fast Fourier Transform and Its Applications, in *Algorithms and Theory of Computation Handbook*, (Second Edition), Volume 1: General Concepts and Techniques, 1016 pp., pages 1–31 in Chapter 18, (Mikhail J. Atallah and Marina Blanton, editors), CRC Press Inc., Boca Raton, Florida, 2010.

[F07] M. Fürer, Faster Integer Multiplication, *Proceedings of 39th Annual Symposium on Theory of Computing (STOC 2007)*, 57–66, ACM Press, New York, 2007.

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

[G98] M. Gu, Stable and Efficient Algorithms for Structured Systems of Linear Equations, *SIAM J. on Matrix Analysis and Applications*, **19**, 279–306, 1998.

[GK72] I. Gohberg, N. Y. Krupnick, A Formula for the Inversion of Finite Toeplitz Matrices, *Matematicheskiie Issledovaniia* (in Russian), **7**, **2**, 272–283, 1972.

[GKO95] I. Gohberg, T. Kailath, V. Olshevsky, Fast Gaussian Elimination with Partial Pivoting for Matrices with Displacement Structure, *Math. of Comp.,* **64(212)**, 1557–1576, 1995.

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

[GO94] I. Gohberg, V. Olshevsky, Complexity of Multiplication with Vectors for Structured Matrices, *Linear Algebra and Its Applications*, **202**, 163–192, 1994.

[GOS08] S. A. Goreinov, I. V. Oseledets, D. V. Savostyanov, E. E. Tyrtyshnikov, N. L. Zamarashkin, How to Find a Good Submatrix, Research Report 08-10, ICM HKBU, Kowloon Tong, Hong Kong, 2008.

[GS72] I. Gohberg, A. Semencul, On the Inversion of Finite Toeplitz Matrices and Their Continuous Analogs, *Matematicheskiie Issledovaniia* (in Russian), **7**, **2**, 187–224, 1972.

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

[GTZ97] S. A. Goreinov, E. E. Tyrtyshnikov, N. L. Zamarashkin, A Theory of Pseudo-skeleton Approximations, *Linear Algebra and Its Applications*, **261**, 1–22, 1997.

[H79] G. Heinig, Beitrage zur spektraltheorie von Operatorbuschen und zur algebraischen Theorie von Toeplitzmatrizen, Dissertation **B**, *TH Karl-Marx-Stadt*, 1979.

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

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

[HR84] G. Heinig, K. Rost, *Algebraic Methods for Toeplitz-like Matrices and Operators, Operator Theory*, **13**, Birkhäuser, 1984.

[K04] I. Kaporin, The Aggregation and Cancellation Techniques As a Practical Tool for Faster Matrix Multiplication, *Theoretical Computer Science*, **315**, **2–3**, 469–510, 2004.

[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.

[KV99] P. Kravanja, M. Van Barel, Algorithms for Solving Rational Interpolation Problems Related to Fast and Superfast Solvers for Toeplitz Systems, *SPIE*, 359–370, 1999.

[LDB02] X. Li, J. Demmel, D. Bailey, G. Henry, Y. Hida, J. Iskandar, W. Kahan, S. Kang, A. Kapur, M. Martin, B. Thompson, T. Tung, D. Yoo, Design, Implementation and Testing of Extended and Mixed Precision BLAS, *ACM Transactions on Mathematical Software*, **28**, 152–205, 2002. http //crd.lbl.gov/x̃iaoye/XBLAS/.

[LPS92] J. Laderman, V. Y. Pan, H. X. Sha, On Practical Algorithms for Accelerated Matrix Multiplication, *Linear Algebra and Its Applications*, **162–164**, 557–588, 1992.

[M80] M. Morf, Doubling Algorithms for Toeplitz and Related Equations, *Proceedings of IEEE International Conference on ASSP*, 954–959, IEEE Press, Piscataway, New Jersey, 1980.

[MMD08] M. W. Mahoney, M. Maggioni, P. Drineas, Tensor-CUR decompositions for tensor-based data, *SIAM Journal on Matrix Analysis and Applications*, **30, 2**, 957–987, 2008.

[MP00] B. Mourrain, V. Y. Pan, Multivariate Polynomials, Duality and Structured Matrices, *J. of Complexity*, **16**, **1**, 110–180, 2000. (Proceedings Version in *STOC'98*.)

[MPR03] B. Mourrain, V. Y. Pan, O. Ruatta, Accelerated Solution of Multivariate Polynomial Systems of Equations, *SIAM J. on Computing*, **32, 2**, 435–454, 2003.

[OOT06] V. Olshevsky, I. V. Oseledets, E. E. Tyrtyshnikov, Tensor Properties of Multilevel Toeplitz and Related Matrices, *Linear Algebra and Its Applications*, **412**, 1–21, 2006.

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

[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.

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

[P90] V. Y. Pan, On Computations with Dense Structured Matrices, *Math. of Computation*, **55, 191**, 179–190, 1990. Also in *Proceedings of International Symposium on Symbolic and Algebraic Computation (ISSAC'89)*, 34–42, ACM Press, NY, 1989.

[P92] V. Y. Pan, Parallel Solution of Toeplitz-like Linear Systems, *J. of Complexity*, **8**, 1–21, 1992.

[P93] V. Y. Pan, Concurrent Iterative Algorithm for Toepliz-like Linear Systems, *IEEE Transactions on Parallel and Distributed Systems*, **4, 5**, 592–600, 1993.

[P93a] V. Y. Pan, Decreasing the Displacement Rank of a Matrix, *SIAM Journal on Matrix Analysis and Applications*, **14**, **1**, 118–121, 1993.

[P98/01] V. Y. Pan, Numerical Computation of a Polynomial GCD and Extensions, *Information and Computation*, **167, 2**, 71–85, 2001. Proc. version: "Approximate Polynomial Gcds, Pad Approximation, Polynomial Zeros, and Bipartite Graphs", in Proc. *9th Ann. ACM-SIAM Symp. on Discrete Algorithms (SODA 98)*, 68-77, ACM Press, New York, and SIAM Publications, Philadelphia (1998).

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

[P09/11] V. Y. Pan, Symbolic Lifting for Structured Linear Systems of Equations, Tech. Report TR 2011002, *Ph.D. Program in Computer Science*, *Graduate Center, the City University of New York*, 2011.
Available at http://www.cs.gc.cuny.edu/tr/techreport.php?id=352

Proc. version in International Symposium on Symbolic-Numerical Computations (Kyoto, Japan, August 2009), (edited by Hiroshi Kai and Hiroshi Sekigawa), pp.105-113, ACM Press, New York (2009).

[Pa] V. Y. Pan, Nearly Optimal Solution of a Linear System of Equations with Preprocessing, preprint, 2011.

[P10] F. Poloni, A Note on the O(n)-Storage Implementation of the GKO Algorithm, *Numerical Algorithms*, **55**, 115–139, 2010.

[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.

[PKRK06] V. Y. Pan, M. Kunin, R. Rosholt, H. Kodal, Homotopic Residual Correction Processes, *Math. of Computation*, **75**, 345–368, 2006.

[PMQR09] V. Y. Pan, B. Murphy, G. Qian, R. E. Rosholt, Error-free Computations via Floating-Point Operations, *Computers and Mathematics with Applications*, **57**, 560–564, 2009.

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

[PQa] V. Y. Pan, G. Qian, Solving Linear System with Randomized Augmentation, Tech. Report TR 2010009, *Ph.D. Program in Computer Science*, *Graduate Center, the City University of New York*, 2010.
Available at http://www.cs.gc.cuny.edu/tr/techreport.php?id=352

[PQZa] V. Y. Pan, G. Qian, A. Zheng, Randomized Preprocessing versus Pivoting, *Linear Algebra and Its Applications*, in print. Also Tech. Report TR 2010011, *Ph.D. Program in Computer Science*, *Graduate Center, the City University of New York*, 2010.

Available at http://www.cs.gc.cuny.edu/tr/techreport.php?id=352.

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

[PS91] V. Y. Pan, R. Schreiber, An Improved Newton Iteration for the Generalized Inverse of a Matrix, with Applications, *SIAM Journal on Scientific and Statistical Computing*, **12, 5**, 1109–1131, 1991.

[PVW04] V. Y. Pan, M. Van Barel, X. Wang, G. Codevico, Iterative Inversion of Structured Matrices, *Theoretical Computer Science*, 315, **2–3** (Special Issue on Algebraic and Numerical Computing, edited by I. Z. Emiris, B. Mourrain, and V. Y. Pan), 581–592, 2004.

[PW08] V. Y. Pan, X. Wang, Degeneration of Integer Matrices Modulo an Integer, *Linear Algebra and Its Applications*, **429**, 2113–2130, 2008.

[PY99/01] V. Y. Pan, Y. Yu, Certified Computation of the Sign of a Matrix Determinant, *Algorithmica*, **30**, 708–724, 2001; Proc. version in *Proc. 10th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'99)*, 715–724, ACM Press, New York, and SIAM Publications, Philadelphia, 1999.

[PZ11] V. Y. Pan, A. Zheng, New Progress in Real and Complex Polynomial Root-Finding, *Computers and Math. (with Applications)* **61**, 1305–1334, 2011.

[R06] G. Rodriguez, Fast Solution of Toeplitz- and Cauchy-like Least Squares Problems, *SIAM J. Matrix Analysis and Applications*, **28, 3**, 724–748, 2006.

[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, Philadelphia, 1998.

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

[ST02] D. Spielman, S.-H. Teng, Smoothed Analysis of Algorithms, *Proc. of the International Congress of Mathematicians* (Beijing 2002), Vol. I, 597–606, Higher ED. Press, Beijing, 2002.

[T90] W. F. Trench, A Note on a Toeplitz Inversion Formula, *Linear Algebra and Its Applications*, **29**, 55–61, 1990.

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

[TB97] L. N. Trefethen, D. Bau III, *Numerical Linear Algebra*, SIAM, Philadelphia, 1997.

[V99] M. Van Barel, A Superfast Toeplitz Solver, 1999.
Available at http://www.cs.kuleuven.be/˜ marc/software/index.html

[VBHK01] M. Van Barel, G. Heinig, P. Kravanja, A Stabilized Superfast Solver for Nonsymmetric Toeplitz Systems, *SIAM Journal on Matrix Analysis and Applications*, **23**, **2**, 494–510, 2001.

[VK98] M. Van Barel, P. Kravanja, A Stabilized Superfast Solver for Indefinite Hankel Systems, *Linear Algebra and its Applications*, **284**, **1–3**, 335–355, 1998.

[VVM07] R. Vandebril, M. Van Barel, N. Mastronardi, *Matrix Computations and Semiseparable Matrices: Linear Systems* (Volume 1), The Johns Hopkins University Press, Baltimore, Maryland, 2007.

[VVM08] R. Vandebril, M. Van Barel, N. Mastronardi, *Matrix Computations and Semiseparable Matrices: Eigenvalue and Singular Value Methods* (Volume 2), The Johns Hopkins University Press, Baltimore, Maryland, 2008.

[W04] M. Wschebor, Smoothed Analysis of $\kappa(a)$, *J. of Complexity*, **20**, 97–107, 2004.

[W07] X. Wang, Affect of Small Rank Modification on the Condition Number of a Matrix, *Computer and Math. (with Applications)*, **54**, 819–825, 2007.

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