Linear randomizing transformations: an a priori alternative to dynamic elimination strategies such as partial pivoting

Brad Pierce
pierce@cs.ucla.edu

D. Stott Parker
stott@cs.ucla.edu

Computer Science Department, UCLA School of Engineering and Applied Science
Los Angeles, CA 90095-1596, USA

December 1996

Abstract

Recently we have reported on new *LU* factorization and block matrix methods that use randomized linear preconditioning in place of dynamic strategies such as partial and pairwise pivoting. We elaborate on the theory underlying these a priori methods, and apply it to several examples, including a new method for the calculation of Moore-Penrose pseudoinverses.

Contents

1 Introduction **2**

2 Preliminaries **3**
2.1 Sequences and submatrices . . . 3
2.2 Rank . . . 4
2.3 Cauchy-Binet formula . . . 4
2.4 Randomness and the zeros of polynomials . . . 5
2.5 Determinants of submatrices of random matrices . . . 6

3 Linear independence: $\Lambda_k^n(\cdot)$ **7**

4 Applications **8**
4.1 The Randomizing Discrete Fourier Transform (RDFT) . . . 8
4.2 Computing pseudoinverses . . . 10
4.2.1 The Moore-Penrose pseudoinverse . . . 10
4.2.2 Generalized LU factorizations . . . 10
4.2.3 A new algorithm for computing pseudoinverses . . . 11

5 Conclusion **12**

References **12**

1 Introduction

The naive approach to the solution of $Ax = b$, where A is a nonsingular $n \times n$ matrix, is the inversion (in n^3 flops) of A, then the multiplication (in n^2 flops) of $A^{-1}b$. More practical is triangularization, *i.e.*, the reduction (in $\frac{1}{3}n^3$ flops) of $Ax = b$ to an equivalent system $(MA)x = Mb$, such that MA is upper triangular, then the solution with back substitution (in $\frac{1}{2}n^2$ flops) of this reduced system.

In **Gaussian elimination**, M and the transformations from which it is constructed are unit lower triangular:

$$\begin{array}{l} J \leftarrow (\, A \quad | \quad b\,) \\ \textbf{for } k \leftarrow 1 \textbf{ to } (n-1) \\ \qquad \textbf{forall } i \in \{(k+1), \ldots, n\} \\ \qquad\qquad J_i \leftarrow J_i - \dfrac{(J_i)_k}{(J_k)_k} J_k \qquad\qquad \{J_i \text{ denotes } i\text{th row}\} \end{array}$$

Observation 1 Gaussian elimination can be successfully applied to an $n \times n$ matrix A iff, for all $1 \leq k \leq n$, the leading submatrix $A[\,1..k \,|\, 1..k\,]$ is nonsingular. (See, for example, [8].)

To handle inputs, *e.g.*, $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, with singular leading submatrices, Gaussian elimination is usually augmented with **partial pivoting**, in which M is the column permutation of a unit lower triangular matrix:

$$\begin{array}{l} J \leftarrow (\, A \quad | \quad b\,) \\ \textbf{for } k \leftarrow 1 \textbf{ to } n \\ \qquad \textbf{let } q \in \{k, \ldots, n\} \textbf{ such that } (J_q)_k \neq 0 \\ \qquad\qquad\qquad J_k \leftrightarrow J_q \\ \qquad \textbf{forall } i \in \{(k+1), \ldots, n\} \\ \qquad\qquad\qquad J_i \leftarrow J_i - \dfrac{(J_i)_k}{(J_k)_k} J_k \quad , \end{array}$$

When A is ill-conditioned, partial pivoting can also be used to improve numerical stability by always choosing a maximal $(J_q)_k$:

$$\textbf{let } q \in \{k, \ldots, n\} \textbf{ such that } \forall p \in \{k, \ldots, n\} \,.\, |(J_q)_k| \;\geq\; |(J_p)_k|$$

This strategy attempts to minimize distortions by making the kth iteration more nearly like transformation by the identity matrix. Although the naive approach in such cases would be to use the theoretically superior method of **Householder reduction** (QR factorization), in which M and the transformations from which it is constructed are unitary, decades of experience have shown that the cheaper method of Gaussian elimination with partial pivoting is usually stable enough.

The advent of modern high-performance computers has largely deprived such comparisons of practical significance, however, since both Householder reduction and partial pivoting require numerous nonlocal data movements, hence are too expensive when implemented on a parallel machine or as a systolic array [6, 10].

The modern context requires methods such as **pairwise pivoting**, in which the computational steps of traditional Gaussian elmination are rearranged to accommodate a local strategy[1]:

[1] As Gentleman and Kung [3, p.22] note, this algorithm was originally used "for the purpose of minimizing storage requirements rather than for avoiding global communication". For example, [2, §13]: "There is one variant which is of particular advantage when the system of equations is fed in as data and not produced by the computer itself, since it enables the solution to be effected using approximately $\frac{1}{2}n^2$ locations instead of n^2."

$J \leftarrow (\, A \quad | \quad b\,)$
for $k \leftarrow 1$ **to** n
  for $j \leftarrow 1$ **to** $(k-1)$
    let $q \in \{k, j\}$ **such that** $\forall p \in \{k, j\}\,.\, |(J_q)_j| \geq |(J_p)_j|$
      $J_j \leftrightarrow J_q$
    if $(J_j)_j \neq 0$ **then**
      $$J_k \leftarrow J_k - \frac{(J_k)_j}{(J_j)_j} J_j$$

Sorensen [13, §4], citing private communication with Wilkinson, Sameh and others, reports that "no adverse effects of rounding error have been observed" in practice. He gives a formal error analysis, but notes that it is "very pessimistic and such growth has not been observed in practice".

This paper explores the theory underlying a new alternative to pairwise pivoting for problems that are known to be well-conditioned, *i.e.*, for the many cases in which instability is not a serious concern.[2] In this approach, first presented in [9], linear systems are preconditioned with presumptive transformations, *e.g.*, application of the FFT to a random reorientation of rows[3] (§4.1), to obtain an equivalent system that is guaranteed to meet the requirements of Observation 1.

Although this strategy is somewhat less efficient than pairwise pivoting, the theory (§3) supporting it is more general. Our approach not only can be applied to many block algorithms [5, 12], but also yields a simple new algorithm (§4.2.3) for the Moore-Penrose pseudoinverse.

2 Preliminaries

2.1 Sequences and submatrices

Definition 1 S_k^n is the set of **strictly increasing sequences** of length k drawn from $\{1, \ldots, n\}$.

For example, $(2,3,5) \in S_3^6$, but $(5,3,2) \notin S_3^6$ and $(2,2,3) \notin S_3^6$. Sequences will be denoted by the initial Greek letters, that is, α, β, and so on. We abbreviate sequences of the form $(1, 2, \ldots, k)$ to $(1..k)$.

Definition 2 Let A be an $n \times p$ matrix, and let $(\alpha, \beta) \in S_{k_1}^n \times S_{k_2}^n$. The **submatrix** $A[\,\alpha\,|\,\beta\,]$ denotes the matrix formed from the entries of the indicated rows and columns of A, that is, if $A = (a_{ij})$ and $A[\,\alpha\,|\,\beta\,] = (b_{ij})$, then $b_{ij} = a_{\alpha_i \beta_j}$.

For example,

$$\begin{pmatrix} 2 & 3 & 5 \\ 7 & 11 & 13 \\ 17 & 19 & 23 \end{pmatrix} [\,1,2\,|\,1,3\,] \quad = \quad \begin{pmatrix} 2 & 5 \\ 7 & 13 \end{pmatrix}.$$

Definition 3 Let A be an $n \times m$ matrix and let $\beta \in S_k^n$ and $\alpha \in S_k^m$. Then $A[\,\vdots\,|\,\alpha\,]$ and $A[\,\beta\,|\,\cdots]$ denote $A[\,1..n\,|\,\alpha\,]$ and $A[\,\beta\,|\,1..m]$, respectively.

[2] For mildly ill-conditioned problems, these methods can be combined with iterative refinement or higher precision.

[3] By 'random reorientation of rows' we mean multiplication of A and b by a diagonal matrix whose nonzero entries are drawn randomly from the complex unit circle.

For example, let X be a 4×6 matrix and partition it as follows:

$$\begin{array}{cc} & 6 \\ 4 & \begin{pmatrix} X \end{pmatrix} \end{array} \quad = \quad \begin{array}{cc} & \begin{array}{cc} 2 & \quad 4 \end{array} \\ \begin{array}{c} 2 \\ 2 \end{array} & \begin{pmatrix} \boxed{A} & \boxed{\quad B \quad} \\ \boxed{C} & \boxed{\quad D \quad} \end{pmatrix} \end{array}$$

then

$$X[\,\vdots\,|\,3,4,5,6\,] \quad = \quad \begin{pmatrix} \boxed{\quad B \quad} \\ \boxed{\quad D \quad} \end{pmatrix}$$

$$X[\,3,4\,|\,\cdots\,] \quad = \quad \begin{pmatrix} \boxed{C} & \boxed{\quad D \quad} \end{pmatrix}$$

2.2 Rank

Definition 4 The **rank** of an $n \times p$ matrix A is the largest $0 \le r \le \min(n,p)$ for which there exists $(\beta, \alpha) \in S_r^n \times S_r^p$, such that $A[\,\beta\,|\,\alpha\,]$ is nonsingular. If $r = \min(n,p)$, then A has **full rank**.

Observation 2 If A has rank r, then, for all $0 \le k \le r$, there exists $(\beta, \alpha) \in S_k^n \times S_k^p$ such that $A[\,\beta\,|\,\alpha\,]$ is nonsingular.[4]

2.3 Cauchy-Binet formula

It is well-known that the determinant of the product of two square matrices is the product of their determinants, *i.e.*, $\det(XY) = (\det X)(\det Y)$. The Cauchy-Binet formula generalizes this product rule to nonsquare matrices:

$$\begin{array}{cc} & p \\ n & \begin{pmatrix} \boxed{\qquad X \qquad} \end{pmatrix} \end{array} \quad \begin{array}{cc} & n \\ p & \begin{pmatrix} \boxed{\begin{array}{c} \\ \\ Y \\ \\ \\ \end{array}} \end{pmatrix} \end{array} \quad = \quad \begin{array}{cc} & n \\ n & \begin{pmatrix} \boxed{XY} \end{pmatrix} \end{array}$$

Lemma 1 *Let X be an $n \times p$ matrix and Y be a $p \times n$ matrix with $n \le p$. Then*

$$\det(XY) \quad = \quad \sum_{\kappa \in S_n^p} \det(X[\,\vdots\,|\,\kappa\,]) \;\; \det(Y[\,\kappa\,|\,\cdots\,])$$

Proof
See, for example, Zeng [14, §2]. □

The Cauchy-Binet formula is often stated in terms of the square submatrices of the product XY, where X and Y are $n \times q$ and $q \times p$ matrices, respectively:

$$\begin{array}{cc} & q \\ n & \begin{pmatrix} \boxed{\qquad X \qquad} \end{pmatrix} \end{array} \quad \begin{array}{cc} & p \\ q & \begin{pmatrix} \boxed{\begin{array}{c} \\ \\ \qquad Y \qquad \\ \\ \\ \end{array}} \end{pmatrix} \end{array} \quad = \quad \begin{array}{cc} & p \\ n & \begin{pmatrix} \boxed{\qquad XY \qquad} \end{pmatrix} \end{array}$$

[4]By Cramer's rule, at least one of the minors of the submatrix in the definition must be nonsingular, as must be at least one of that minor's minors, and so on.

Theorem 1 *Let X be an $n \times q$ matrix and Y be a $q \times p$ matrix. If $0 \le k \le \min(n,q,p)$ and $(\beta,\alpha) \in S_k^n \times S_k^p$, then*

$$\det\left((XY)[\,\beta\,|\,\alpha\,]\right) \quad = \quad \sum_{\kappa \in S_k^q} \det(X[\,\beta\,|\,\kappa\,])\ \det(Y[\,\kappa\,|\,\alpha\,]) \qquad .$$

Proof

Immediate from Lemma 1. Specifically, it follows from the definition of matrix multiplication

$$(XY)[\,i\,|\,j\,] \quad = \quad X[\,i\,|\,\cdots]\,Y[\,\vdots\,|\,j\,]$$

that for $(\beta,\alpha) \in S_k^n \times S_k^p$, with $k < \min(n,q,p)$,

$$(XY)[\,\beta\,|\,\alpha\,] \quad = \quad X[\,\beta\,|\,\cdots] \quad Y[\,\vdots\,|\,\alpha\,]$$

hence

$$\begin{aligned}
\det((XY)[\,\beta\,|\,\alpha\,]) &= \det(\ X[\,\beta\,|\,\cdots]\,Y[\,\vdots\,|\,\alpha\,]\ ) \\
&= \sum_{\kappa \in S_k^q} \det((X[\,\beta\,|\,\cdots])[\,\vdots\,|\,\kappa\,])\ \det((Y[\,\vdots\,|\,\alpha\,])[\,\kappa\,|\,\cdots]) \\
&= \sum_{\kappa \in S_k^q} \det(X[\,\beta\,|\,\kappa\,])\ \det(Y[\,\kappa\,|\,\alpha\,])
\end{aligned}$$

□

2.4 Randomness and the zeros of polynomials

Definition 5 A **random real index** θ is a real-valued integrable function $\theta : \mathcal{D} \to \mathbf{R}$ with a strictly increasing (absolutely) continuous probability distribution on a real interval sample space $\mathcal{D}$ of nonzero measure.

A 1-to-1 complex-valued function of a random real index θ is a **random value** if it is integrable with respect to the probability distribution of θ. A **nonzero random value** is a random value that is zero almost never.

Definition 6 A **random matrix** is a matrix whose complex-valued entries are functions of independent random values sharing a common distribution.

An $n \times n$ **random diagonal matrix** is a random matrix of the form

$$\mathrm{diag}(r_1, r_2, \ldots, r_n)\ ,$$

where $r_1, r_2, \ldots, r_n$ are independently chosen random values.

An $n \times n$ **unitary random diagonal matrix** is a random diagonal matrix of the form

$$\mathrm{diag}(e^{\mathrm{i}\theta_1}, \ldots, e^{\mathrm{i}\theta_n})\ ,$$

where $\theta_1, \ldots, \theta_n$ are independently chosen random real indexes.

Observation 3 If R is a unitary random diagonal matrix, then so is R^*.

Definition 7 If $r_1, r_2, \ldots, r_n$ are a random values, a property $P(r_1, r_2, \ldots, r_n)$ holds **with probability 1** when $P(r_1, r_2, \ldots, r_n)$ is true on all of $\mathcal{D}^n$ except possibly a subset of measure 0.[5] Equivalently, it holds **almost always**. A property $P(r_1, r_2, \ldots, r_n)$ holds **almost never** if $\neg P(r_1, r_2, \ldots, r_n)$ holds almost always.

Theorem 2 *If $p(x_1, \ldots, x_n)$ is a nonzero polynomial with complex coefficients and $r_1, \ldots, r_n$ are independently chosen random values, then $p(r_1, \ldots, r_n) \neq 0$ almost always.*

Proof
p has a finite number of roots, each of them an n-tuple from $\mathbf{C}^n$. Because the random values are chosen independently, $(r_1, \ldots, r_n)$ is one of these roots almost never.

2.5 Determinants of submatrices of random matrices

Observation 4 If $\widetilde{H}$ is an $n \times q$ random matrix and, for some $1 \leq k \leq \min(n, q)$, $(\beta, \kappa) \in S_k^n \times S_k^p$, then $\det \widetilde{H}[\,\beta\,|\,\kappa\,]$ is a random function.

For example, if

$$\widetilde{H} \;=\; \begin{pmatrix} r & 1 & s \\ 0 & 1 & 0 \\ t & 1 & v \end{pmatrix}$$

for independently chosen random values r, s, t, v, then

$$\begin{aligned} \det \widetilde{H}[\,1,3\,|\,2,3\,] &= \det \begin{pmatrix} 1 & s \\ 1 & v \end{pmatrix} \\ &= v - s \end{aligned}$$

Because v and s are drawn from the same distribution (see Definition 6), the expected value of $\det \widetilde{H}[\,1,3\,|\,2,3\,]$ is zero. Nevertheless, by Theorem 2, $\widetilde{H}[\,1,3\,|\,2,3\,]$ is nonsingular almost always.

Observation 5 Let $D = \operatorname{diag}(d_1, d_2, \ldots, d_n)$. If $\beta \in S_k^n$ with $1 \leq k \leq n$, then

$$\det D[\,\beta\,|\,\beta\,] \;=\; \prod_{i=1}^{k} d_{\beta_i} \quad .$$

If $\alpha \in S_k^n$ with $\beta \neq \alpha$, however, then $\det D[\,\beta\,|\,\alpha\,] = \det D[\,\alpha\,|\,\beta\,] = 0$.

Theorem 3 *Let $\widetilde{U}$ be an $n \times q$ random matrix, R be a $q \times q$ random diagonal matrix, and $\widetilde{V}$ be a $q \times p$ random matrix.*[6] *Let $(\beta, \alpha) \in S_k^n \times S_k^p$ with $0 \leq k \leq \min(n, q, p)$. If almost always there exists a $\kappa \in S_k^q$ such that both $\widetilde{U}[\,\beta\,|\,\kappa\,]$ and $\widetilde{V}[\,\kappa\,|\,\alpha\,]$ are nonsingular, then $(\widetilde{U}R\widetilde{V})[\,\beta\,|\,\alpha\,]$ is nonsingular almost always.*

[5] To support implementation with floating point numbers the quantity n of random numbers used must be 'small' relative to the number of floating point numbers available. If there are 2^w possible floating point numbers, $n \ll 2^w$ is necessary for the matrices to be considered random in the sense intended here. Furthermore, the phrase "with probability 1" must be understood as relative to floating point numbers. If floating point numbers have t-bit mantissa precision, then (ignoring exponents and normalization and distribution) the probability that two arbitrarily-chosen floating point numbers are equal is at least 2^{-t}, and (assuming a good pseudorandom number generator) the phrase must be replaced by "with probability $1 - O(2^{-t})$". Since these $O(2^{-t})$ terms are distracting and are not likely to be significant (most computers today support $t \geq 48$), we omit them from now on, but the phrase "with probability 1" should be treated with care.

[6] Assume, moreover, that $\widetilde{U}$, R, and $\widetilde{V}$ are independent of each other.

Proof
Let $R = \mathrm{diag}(r_1, r_2, \ldots, r_q)$. By the Cauchy-Binet formula

$$\begin{aligned}\det((\widetilde{U}R\widetilde{V})[\,\beta\,|\,\alpha\,]) &= \sum_{\kappa \in S_k^q} \det \widetilde{U}[\,\beta\,|\,\kappa\,]\ \det R[\,\kappa\,|\,\kappa\,]\ \det \widetilde{V}[\,\kappa\,|\,\alpha\,] \\ &= \sum_{\kappa \in S_k^q} \det \widetilde{U}[\,\beta\,|\,\kappa\,] \left(\prod_{i=1}^{k} r_{\kappa_i}\right) \det \widetilde{V}[\,\kappa\,|\,\alpha\,]\end{aligned}$$

This is a polynomial in the random values of R with each index κ corresponding to a monomial summand constructed from a coefficient $\det \widetilde{U}[\,\beta\,|\,\kappa\,] \det \widetilde{V}[\,\kappa\,|\,\alpha\,]$ and a distinct combination of random values. Moreover, because almost always there exists a $\kappa \in S_k^q$ such that both $\widetilde{U}[\,\beta\,|\,\kappa\,]$ and $\widetilde{V}[\,\kappa\,|\,\alpha\,]$ are nonsingular, this polynomial is nontrivial almost always. The conclusion follows from Theorem 2. □

3 Linear independence: $\Lambda_k^n(\cdot)$

Let A be a nonsingular $n \times n$ matrix and $\widetilde{U}$ be an $n \times n$ random matrix. Consider $(\widetilde{U}A)[\,\beta\,|\,\alpha\,]$ for some $(\beta, \alpha) \in S_k^n \times S_k^n$. By the Cauchy-Binet formula

$$\det(\widetilde{U}A)[\,\beta\,|\,\alpha\,] \quad = \quad \sum_{\kappa \in S_k^n} \det A[\,\kappa\,|\,\alpha\,]\ \det \widetilde{U}[\,\beta\,|\,\kappa\,]\ .$$

Because A is fixed, each of the $A[\,\kappa\,|\,\alpha\,]$ are constants, hence this sum is just a linear combination of the random functions

$$\left\{ \det \widetilde{U}[\,\beta\,|\,\kappa\,] \quad | \quad \kappa \in S_k^q \right\} \quad .$$

Moreover, because A is nonsingular, there exists a $\kappa \in S_k^n$ such that $\det A[\,\kappa\,|\,\alpha\,] \neq 0$, hence, the linear combination is nontrivial.

If the random functions were linearly independent, *i.e.*, if every nontrivial linear combination almost never were zero, then $(\widetilde{U}A)[\,\beta\,|\,\alpha\,]$ would almost always be nonsingular.

Definition 8 Let $\widetilde{H}$ be an $n \times q$ random matrix. $\Lambda_k^n(\widetilde{H}) \subseteq S_k^n$ with $1 \leq k \leq \min(n, q)$ consists of those $\beta \in S_k^n$, such that all nontrivial linear combinations of the random functions

$$\left\{ \det \widetilde{H}[\,\beta\,|\,\kappa\,] \quad | \quad \kappa \in S_k^q \right\}$$

almost never are zero, that is, such that, for all $\psi : S_k^n \to \mathbf{C}$ with $\psi(S_k^n) \neq \{0\}$,

$$\sum_{\kappa \in S_k^n} \psi(\kappa)\ \det \widetilde{H}[\,\beta\,|\,\kappa\,]$$

almost never is zero.

Observation 6 $\Lambda_k^q(\widetilde{H}^*)$ consists of those $\alpha \in S_k^q$, such that all nontrivial linear combinations of the random functions

$$\left\{ \det \widetilde{H}[\,\kappa\,|\,\alpha\,] \quad | \quad \kappa \in S_k^n \right\}$$

almost never are zero.

Theorem 4 *Let $\widetilde{U}$ be an $n \times q$ random matrix, A be a $q \times p$ matrix, and $0 \le k \le \min(n, q, p)$. Let $(\beta, \alpha) \in \Lambda_k^n(\widetilde{U}) \times S_k^p$. If $A[\,\vdots \,|\, \alpha\,]$ has full rank, then $(\widetilde{U}A)[\,\beta \,|\, \alpha\,]$ is nonsingular almost always.*

Proof
By the Cauchy-Binet formula

$$\det(\widetilde{U}A)[\,\beta \,|\, \alpha\,] \;\; = \;\; \sum_{\kappa \in S_k^q} \det A[\,\kappa \,|\, \alpha\,] \;\det \widetilde{U}[\,\beta \,|\, \kappa\,] \;.$$

Because $A[\,\vdots \,|\, \alpha\,]$ has full rank, there exists a $\kappa \in S_k^q$ such that $A[\,\kappa \,|\, \alpha\,]$ is nonsingular, hence such that $\det A[\,\kappa \,|\, \alpha\,] \neq 0$. Because $\beta \in \Lambda_k^n(\widetilde{U})$, this nontrivial linear combination is zero almost never. □

Observation 7 If A is a nonsingular $n \times n$ matrix, and $\widetilde{U}$ is an $n \times n$ random matrix, such that, for all $1 \le k \le n$, $(1..k) \in \Lambda_k^n(\widetilde{U})$, then Gaussian elimination (without any pivoting) can be successfully applied to $\widetilde{U}A$ almost always. Gaussian elimination cannot necessarily be used to solve the problem $Ax = b$, because A may have singular leading submatrices, but if U is an instance of $\widetilde{U}$, then Gaussian elimination can be used to solve the equivalent problem $(UA)x = Ub$ almost always.

Corollary 1 *Let A be a $q \times p$ matrix, $\widetilde{V}$ be an $m \times p$ random matrix, and $0 \le k \le \min(q, p, m)$. Let $(\beta, \alpha) \in S_k^q \times \Lambda_k^m(\widetilde{V})$. If $A[\,\beta \,|\, \cdots\,]$ has full rank, then $(A\widetilde{V}^*)[\,\beta \,|\, \alpha\,]$ is nonsingular almost always.*

Theorem 5 *Let $\widetilde{U}$ be an $n \times q$ random matrix and $\widetilde{V}$ be an $m \times p$ random matrix.[7] Let A be a $q \times p$ matrix of rank r, and let $0 \le k \le r$. If $(\beta, \alpha) \in \Lambda_k^n(\widetilde{U}) \times \Lambda_k^m(\widetilde{V})$, then $(\widetilde{U}A\widetilde{V}^*)[\,\beta \,|\, \alpha\,]$ is nonsingular almost always.*

Proof
Because A has rank r, there must exist $(\kappa, \gamma) \in S_k^q \times S_k^p$ such that $\det A[\,\kappa \,|\, \gamma\,] \neq 0$, hence such that $A[\,\vdots \,|\, \gamma\,]$ has full rank. By Theorem 4 $(\widetilde{U}A)[\,\beta \,|\, \gamma\,]$ almost always is nonsingular, hence $(\widetilde{U}A)[\,\beta \,|\, \cdots\,]$ almost always has full rank. The conclusion follows from Corollary 1. □

4 Applications

4.1 The Randomizing Discrete Fourier Transform (RDFT)

We introduced the RDFT in [11]. Using $\Lambda_k^n(\cdot)$ and Theorem 4, our earlier results can be more perspicuously stated and proved.

Definition 9 The **Discrete Fourier Transform (DFT)** is the family of unitary $n \times n$ matrices $F_\omega = (f_{ij})$ defined by

$$f_{ij} \;\; = \;\; \frac{1}{\sqrt{n}} \left(\omega^{(i-1)(j-1)} \right),$$

where ω is a principal nth root of unity, that is, $\omega^k = 1$ iff $k = jn$ for some integer j.

[7] Assume, moreover, that $\widetilde{U}$ and $\widetilde{V}$ are independent of each other.

Definition 10 A **Randomizing Discrete Fourier Transformation (RDFT)** (resp. **adjoint RDFT**) is any random matrix of the form $F_\omega R$ (resp. RF_ω^*), where F_ω is a DFT and R is a unitary random diagonal matrix.

Observation 8 Every instance of an RDFT (or an adjoint RDFT) is unitary.

Theorem 6 *Let A be an $n \times p$ matrix and $\beta \in S_k^n$ with $1 \leq k \leq \min(n,p)$. Let R be a $p \times p$ random diagonal matrix. If, for all $\kappa \in S_k^p$, $A[\,\beta\,|\,\kappa\,]$ is nonsingular, then $\beta \in \Lambda_k^n(AR)$.*

Proof
Let $R = \mathrm{diag}(r_1, r_2, \ldots, r_p)$, and let $\psi : S_k^p \to \mathbf{C}$ with $\psi(\widehat{\kappa}) \neq 0$. By the Cauchy-Binet formula

$$\begin{aligned}\sum_{\kappa \in S_k^p} \psi(\kappa) \det(AR)[\,\beta\,|\,\kappa\,] &= \sum_{\kappa \in S_k^p} \psi(\kappa) \sum_{\gamma \in S_k^p} \det A[\,\beta\,|\,\gamma\,] \ \det R[\,\gamma\,|\,\kappa\,] \\ &= \sum_{\kappa \in S_k^p} \psi(\kappa) \ \det A[\,\beta\,|\,\kappa\,] \prod_{i=1}^{k} r_{\kappa_i}\end{aligned}$$

Thus $\sum_{\kappa \in S_k^p} \psi(\kappa) \det(AR)[\,\beta\,|\,\kappa\,]$ is a polynomial in the random values of R with each index κ corresponding to a monomial summand constructed from a coefficient $\psi(\kappa) \det A[\,\beta\,|\,\kappa\,]$ and a distinct combination of random values. Moreover, because $\psi(\widehat{\kappa}) \det A[\,\beta\,|\,\widehat{\kappa}\,]$ almost never is zero, this polynomial almost always is nontrivial. The conclusion follows from Theorem 2. □

Observation 9 If ω is a principal nth root of unity and $0 \leq j < i < n$, then $\omega^i \neq \omega^j$.[8]

Lemma 2 *Let F_ω be a DFT matrix of size n. Then for all $1 \leq k \leq n$ and all sequences $\alpha \in S_k^n$, the submatrix $F_\omega[\,1..k\,|\,\alpha\,]$ is nonsingular.*

Proof
$F_\omega[\,1..k\,|\,\alpha\,]$ is the $k \times k$ matrix

$$\frac{1}{\sqrt{n}} \begin{pmatrix} \omega^{(\alpha_1 - 1)\,0} & \omega^{(\alpha_2 - 1)\,0} & \ldots & \omega^{(\alpha_k - 1)\,0} \\ \omega^{(\alpha_1 - 1)\,1} & \omega^{(\alpha_2 - 1)\,1} & \ldots & \omega^{(\alpha_k - 1)\,1} \\ \vdots & \vdots & \ddots & \vdots \\ \omega^{(\alpha_1 - 1)(k-1)} & \omega^{(\alpha_2 - 1)(k-1)} & \ldots & \omega^{(\alpha_k - 1)(k-1)} \end{pmatrix}.$$

Thus the submatrix is a constant multiple of a transposed Vandermonde matrix (x_j^{i-1}), where $x_j = \omega^{(\alpha_j - 1)}$. Because ω is a principal nth root of unity, the x_j are distinct, hence the Vandermonde matrix must be nonsingular [4, p.178ff]. □

Theorem 7 *If $F_\omega R$ is an RDFT of size n, then for all $1 \leq k \leq n$, $(1..k) \in \Lambda_k^n(F_\omega R)$.*

Proof
Immediate from Lemma 2 and Theorem 6. □

Observation 10 If RF_ω^* is an adjoint RDFT of size n, then for all $1 \leq k \leq n$, $(1..k) \in \Lambda_k^n((RF_\omega^*)^*)$.

[8] Assume for contradiction that $\omega^i = \omega^j$, then $1 = \omega^{ni} = \omega^i \omega^{(n-1)i} = \omega^i \omega^{(n-1)j} = \omega^{(i-j)+nj} = \omega^{i-j}$.

The following result is new:

Theorem 8 *Let F_{ω_1} and F_{ω_2} be DFT matrices of size n, and let R_1 be an $n \times n$ random diagonal matrix. Then, for all $(\beta, \alpha) \in S_k^n \times S_k^n$ with $1 \leq k \leq n$, $(F_{\omega_2}^* R_1 F_{\omega_1})[\,\beta \,|\, \alpha\,]$ is nonsingular almost always. Moreover, if R_2 is an $n \times n$ random diagonal matrix (with random values independent of those in R_1) then*

$$\Lambda_k^n(F_{\omega_2}^* R_1 F_{\omega_1} R_2) \quad = \quad S_k^n$$

Proof
By Lemma 2, $F_{\omega_2}^*[\,\beta \,|\, 1..k\,]$ and $F_{\omega_1}[\,1..k \,|\, \alpha\,]$ are nonsingular. The conclusions follow, respectively, from Theorems 3 and 6. □

4.2 Computing pseudoinverses

Using $\Lambda_k^n(\cdot)$ and Theorem 5, we have discovered and proved a simple new algorithm for computing the Moore-Penrose pseudoinverse of an arbitrary rectangular matrix.

4.2.1 The Moore-Penrose pseudoinverse

Definition 11 The *Moore-Penrose pseudoinverse* of an $n \times p$ matrix A is the unique $p \times n$ matrix A^+ that satisfies:

- $AA^+A = A$
- $A^+AA^+ = A^+$
- AA^+ and A^+A are Hermitian

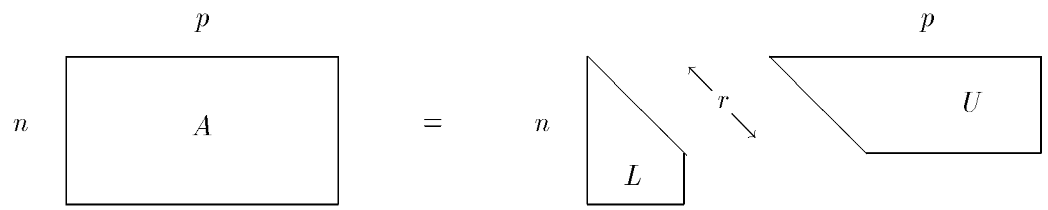
Some important facts about pseudoinverses:

- If A is nonsingular, $A^+ = A^{-1}$.
- If P and Q are unitary, then $(PAQ)^+ = Q^*A^+P^*$.
- If A and B both have full rank, then $(AB)^+ = B^+A^+$.
- If A is an $n \times p$ matrix with full rank, then $A^+ = \begin{cases} (A^*A)^{-1}A^*, & \text{if } n \geq p \\ A^*(AA^*)^{-1}, & \text{if } n \leq p \end{cases}$.

See, for example, [1].

4.2.2 Generalized LU factorizations

Definition 12 Let A be an $n \times p$ matrix with rank r. A **generalized LU factorization** of A is a pair of full rank matrices, L and U, such that:



where $L[\,1..r\,|\,1..r\,]$ is unit lower triangular and $U[\,1..r\,|\,1..r\,]$ is upper triangular.

When A is nonsingular, then $L = M^{-1}$, where M is the transformation performed by Gaussian elimination. Even if Gaussian elimination cannot be applied to A, because a leading submatrix is singular, there is still always a permutation matrix P, such that Gaussian elimination is applicable to PA, hence such that PA has an LU factorization. Constructing such a P is the essence of partial pivoting.

In the general case, both rows and columns must be rearranged. For any matrix A, there are always permutation matrices P and Q such that PAQ has a generalized LU factorization. These permutations can be constructed using complete pivoting.

With randomizing transformations we can instead presumptively generate (nonpermutation) matrices P and Q, such that PAQ has a generalized LU factorization almost always.

If $(PAQ)[\,1..r\,|\,1..r\,]$ has the LU factorization L_rU_r, then the generalized LU factorization of PAQ is

$$L \;=\; \begin{array}{c} \\ r \\ n-r \end{array}\!\!\begin{array}{c} r \\ \begin{pmatrix} L_r \\ CU_r^{-1} \end{pmatrix} \end{array} \qquad\qquad U \;=\; \begin{array}{c} \\ r \end{array}\!\!\begin{array}{cc} r & p-r \\ \multicolumn{2}{c}{\begin{pmatrix} U_r & L_r^{-1}B \end{pmatrix}} \end{array} \quad ,$$

where

$$PAQ \;=\; \begin{array}{c} \\ r \\ n-r \end{array}\!\!\begin{array}{cc} r & p-r \\ \multicolumn{2}{c}{\begin{pmatrix} L_rU_r & B \\ C & D \end{pmatrix}} \end{array} \quad .$$

Hence, if the value of r were unknown, but it were known that, for all $1 \le k \le r$, $(PAQ)[\,1..k\,|\,1..k\,]$ is nonsingular, then the generalized LU factorization of PAQ could be computed by applying Gaussian elimination (without any pivoting) until reaching a zero diagonal entry (or terminating normally):

$(L, U, r) \leftarrow (I_n, PAQ, \min(n, p))$

for $k \leftarrow 1$ **to** $\min(n, p)$
  if $(U_k)_k = 0$ **then** $\{r \leftarrow (k-1)\ ;$ **break**$\}$
  forall $i \in \{(k+1), \ldots, n\}$
    $(L_i)_k \leftarrow \dfrac{(U_i)_k}{(U_k)_k}$
    $U_i \leftarrow U_i - (L_i)_k U_k$

$(L, U) \leftarrow (L[\,1..n\,|\,1..r\,], U[\,1..r\,|\,1..p\,])$

4.2.3 A new algorithm for computing pseudoinverses

Noble [7, §6] describes a "method for finding the generalized inverse, based on Gauss elimination with complete pivoting". Here we modify this method, using two unitary randomizing transforms in place of complete pivoting.

Let A be an $n \times p$ matrix. Let $\widetilde{P}$ be a unitary $n \times n$ random matrix such that, for all $1 \le k \le n$, $(1..k) \in \Lambda_k^n(\widetilde{P})$ and let $\widetilde{Q}$ be a unitary $p \times p$ random matrix such that, for all $1 \le k \le p$, $(1..k) \in \Lambda_k^p(\widetilde{Q})$.

Generate instances P and Q of $\widetilde{P}$ and $\widetilde{Q}$ respectively. Then almost always Gaussian elimination without pivoting can be applied to PAQ^* to calculate a generalized LU factorization, such that

$PAQ^* = LU$, hence such that $A = P^*(LU)Q$. Then

$$\begin{aligned} A^+ &= Q^*(LU)^+(P^*)^* \\ &= Q^*(U^+L^+)P \end{aligned}$$

Note that, if P and Q^* are pseudorandom, then they needn't be saved explicitly. Seeds could be saved instead, and the matrices regenerated when needed.

5 Conclusion

This paper has elaborated on the theory underlying a new approach to linear systems solution, in which problems are randomized with presumptive linear transformations to obtain equivalent problems that are guaranteed to meet the preconditions of a given algorithm.

Specifically, we have introduced the concept of the set of sequences $\Lambda_k^n(\widetilde{H})$, where $\widetilde{H}$ is a random matrix. If $\widetilde{H}$ is $n \times q$, then $\Lambda_k^n(\widetilde{H})$ denotes the set of $\beta \in S_k^n$, such that the random functions

$$\left\{ \det \widetilde{H}[\,\beta \,|\, \kappa\,] \quad | \quad \kappa \in S_k^q \right\}$$

are linearly independent.

Using $\Lambda_k^n(\cdot)$, we have more perspicuously restated earlier results and have discovered a simple new algorithm for computing the Moore-Penrose pseudoinverse.

References

[1] A. Ben-Israel and T.N.E. Greville. *Generalized Inverses: Theory and Applications.* Wiley, NY, 1974.

[2] National Physical Laboratory (Great Britain). *Modern computing methods.* H.M. Stationery Off., London, second edition, 1961.

[3] W.M. Gentleman and H.T. Kung. Matrix triangularization by systolic arrays. *Proceedings of the SPIE - The International Society for Optical Engineering*, 298:19–26, August 1981.

[4] G.H. Golub and C.F. Van Loan. *Matrix Computations.* Johns Hopkins University Press, Baltimore, MD, second edition, 1989.

[5] D. Lê and D.S. Parker. Practical recursive block decomposition matrix algorithms, and quadtree algorithms, via randomization. Technical Report CSD-950039, UCLA Computer Science Department, 1995.

[6] D. Lê, D.S. Parker, and B. Pierce. Input randomization and automatic derivation of systolic arrays for matrix computations. Technical Report CSD-950038, UCLA Computer Science Department, 1995.

[7] B. Noble. Methods for computing the Moore-Penrose generalized inverse, and related matters. In M.Z. Nashed, editor, *Generalized Inverses and Applications : Proceedings of an Advanced Seminar*, pages 245–302. Academic Press, NY, 1976.

[8] D.S. Parker. Explicit formulas for the results of Gaussian elimination. Technical Report CSD-950025, UCLA Computer Science Department, 1995.

[9] D.S. Parker. Random butterfly transformations with applications in computational linear algebra. Technical Report CSD-950023, UCLA Computer Science Department, 1995.

[10] D.S. Parker and D. Lê. How to eliminate pivoting from Gaussian elimination — by randomizing instead. Technical Report CSD-950022, UCLA Computer Science Department, 1995.

[11] D.S. Parker and B. Pierce. The randomizing FFT: an alternative to pivoting in Gaussian elimination. Technical Report CSD-950037, UCLA Computer Science Department, 1995.

[12] B. Pierce and D.S. Parker. A block matrix generalization of Gauss-Jordan elimination using Haynsworth's quotient formula for Schur complements. Technical Report CSD-950063, UCLA Computer Science Department, 1995.

[13] D.C. Sorensen. Analysis of pairwise pivoting in Gaussian elimination. *IEEE Trans. Comput.*, C-34(3):274–278, March 1985.

[14] J. Zeng. A bijective proof of Muir's identity and the Cauchy-Binet formula. *Linear Algebra Appl.*, 184:79–82, 1993.