$\boldsymbol{LU}$ AND $\boldsymbol{CR}$ ELIMINATION

GILBERT STRANG AND CLEVE MOLER

Abstract. The reduced row echelon form $\mathbf{rref}(A)$ has traditionally been used for classroom examples: small matrices A with integer entries and low rank r. This paper creates a column-row rank-revealing factorization $\boldsymbol{A = CR}$, with the first r independent columns of A in C and the r nonzero rows of $\mathbf{rref}(A)$ in R. We want to reimagine the start of a linear algebra course, by helping students to see the independent columns of A and the rank and the column space.

If B contains the first r independent rows of A, then those rows of $\boldsymbol{A = CR}$ produce $\boldsymbol{B = WR}$. The r by r matrix W has full rank r, where B meets C. Then the triple factorization $\boldsymbol{A = CW^{-1}B}$ treats columns and rows of A (C and B) in the same way.

Key words. elimination, factorization, row echelon form, matrix, rank

AMS subject classifications. 15A21, 15A23, 65F05, 65F55

1. Introduction.

Matrix factorizations like $A = LU$ and $A = U\Sigma V^{\mathrm{T}}$ have become the organizing principles of linear algebra. This expository paper develops a column-row factorization $\boldsymbol{A = CR} = (m \times r)\,(r \times n)$ for any matrix of rank r. The matrix C contains the *first r independent columns* of A: a basis for the column space. The matrix R contains the *nonzero rows* of the reduced row echelon form $\mathbf{rref}(A)$. We will put R in the form $\boldsymbol{R} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{F} \end{bmatrix} \boldsymbol{P}$, with an r by r identity matrix that multiplies the r columns of C. Then $A = CR = \begin{bmatrix} C & CF \end{bmatrix} P$ expresses the $n - r$ remaining (dependent) columns of A as combinations CF of the r independent columns in C. When those independent columns don't all come first in A, P permutes those columns of I and F into their correct positions.

The example in Section 3 shows how invertible row operations find the first r independent columns of A. For a large matrix this row reduction is expensive and numerically perilous. But Section 6 will explain the value of an approximate CR or $CW^{-1}B$ factorization of A. This is achievable by randomized linear algebra.

The key point is: $A = CR$ is an "interpolative decomposition" that includes r actual columns of A in C. A more symmetric two-sided factorization $\boldsymbol{A = C\,W^{-1}B}$ also includes r actual rows of A in B. The r by r matrix W lies at the "intersection" inside A of the columns of C with the rows of B. The mixing matrix W^{-1} removes that repetition to produce $\boldsymbol{W^{-1}B = R}$. If $P = I$ then we are seeing block elimination with W as the block pivot:

$$A = \begin{bmatrix} W & H \\ J & K \end{bmatrix} = \begin{bmatrix} \boldsymbol{W} \\ \boldsymbol{J} \end{bmatrix} \boldsymbol{W^{-1}} \begin{bmatrix} \boldsymbol{W} & \boldsymbol{H} \end{bmatrix} = \boldsymbol{C\,W^{-1}B}$$

That matrix W is invertible, where a row basis B meets the column basis C. For large matrices, a low rank version $A \approx CUB$ can give a high quality approximation.

2. *LU* Elimination. This is familiar to all numerical analysts. It applies best to an invertible n by n matrix A. A typical step subtracts a multiple ℓ_{ij} of row j from row i, to produce a zero in the ij position for each $i > j$. A column at a time, going left to right, all entries below the main diagonal become zero. LU elimination arrives at an upper triangular matrix U, with the n nonzero pivots on its main diagonal.

We could express that result as a matrix multiplication $EA = U$. The lower triangular matrix E is the product of all the single-step elimination matrices E_{ij}. A more successful idea—which reverses the order of the steps as it inverts them—is to consider the matrix $L = E^{-1}$ that brings U back to A. Then $EA = U$ becomes $A = LU$.

In this order, the lower triangular L contains all the multipliers ℓ_{ij} *exactly in their proper positions*. The pattern is only upset if any row exchanges become necessary to avoid a zero pivot or to obtain a larger pivot (and smaller multipliers). If all row exchanges are combined into a permutation matrix, elimination factors that row-permuted version of A into LU.

3. *CR* Elimination. Start now with an m by n matrix A of rank r. Elimination will again proceed left to right, a column at a time. The new goal is to produce an r by r identity matrix I. So each pivot row in turn is divided by its first nonzero entry, to produce the desired 1 in I. Then multiples of that pivot row are subtracted from the rows *above and below*, to achieve the zeros in that column of I.

If at any point a row is entirely zero, it moves to the bottom of the matrix. If a column is entirely zero except in the rows already occupied by earlier pivots, then no pivot is available in that column—and we move to the next column. The final result of this elimination is the m by n **reduced row echelon form** of A. We denote that form by R_0 :

$$\boldsymbol{R_0} = \mathbf{rref}(\boldsymbol{A}) = \begin{array}{c} \left[\begin{array}{cc} \boldsymbol{I} & \boldsymbol{F} \\ \mathbf{0} & \mathbf{0} \end{array}\right] \boldsymbol{P} \\ \begin{array}{cc} r & n-r \end{array} \text{ columns} \end{array} \qquad \begin{array}{ll} r & \text{rows} \\ m-r & \text{rows} \end{array} \tag{3.1}$$

The n by n permutation matrix P puts the columns of $I_{r\times r}$ into their correct positions, matching the positions of the first r independent columns of the original matrix A.

$$A = \begin{bmatrix} \mathbf{1} & 2 & 3 & \mathbf{4} \\ \mathbf{1} & 2 & 3 & \mathbf{5} \\ \mathbf{2} & 4 & 6 & \mathbf{9} \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} \mathbf{1} & \mathbf{2} & \mathbf{3} & 0 \\ 0 & 0 & 0 & \mathbf{1} \\ 0 & 0 & 0 & 0 \end{bmatrix} = \boldsymbol{R_0}. \tag{3.2}$$

C contains columns 1 and 4 of A and R contains rows 1 and 2 of R_0. Then $\boldsymbol{CR}$ contains the four columns of A : column 1, 2(column 1), 3(column 1), column 4. P^{T} puts I first :

(3.3)

$$\boldsymbol{R_0 P^{\mathrm{T}}} = \begin{bmatrix} \mathbf{1} & \mathbf{2} & \mathbf{3} & 0 \\ 0 & 0 & 0 & \mathbf{1} \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 \\ 0 & 0 & 0 & \mathbf{1} \\ 0 & \mathbf{1} & 0 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{1} & 0 & \mathbf{2} & \mathbf{3} \\ 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{F} \\ 0 & 0 \end{bmatrix}.$$

All permutations have $P^{\mathrm{T}}P = I$. So multiplying equation (3.3) by P and removing row 3 produces $\boldsymbol{R} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{F} \end{bmatrix} \boldsymbol{P}$.

Usually the description of elimination stops at R_0. There is *no connection* to a matrix factorization $A = CR$. And the matrix F in the columns without pivots is given *no interpretation*. This misses an opportunity to understand more fully the **rref** algorithm and the structure of a general matrix A. We believe that $A = CR$ is a valuable idea in teaching linear algebra [11] : a good way to start.

4. The Factorization $\boldsymbol{A = CR}$: $\boldsymbol{m}$ by $\boldsymbol{r}$ times $\boldsymbol{r}$ by $\boldsymbol{n}$. The matrix C is easy to describe. It contains the *first r independent columns of A*. The positions of those independent columns are revealed by the identity matrix I in $\mathbf{rref}(A)$ and by P (the permutation). All other columns of A (with rank r) are combinations of these r columns. Those combinations come from the submatrix F. **The matrix $\boldsymbol{R}$ is the reduced row echelon form $\boldsymbol{R_0} = \mathbf{rref}(\boldsymbol{A})$ without its zero rows**:

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{F} \end{bmatrix} \boldsymbol{P} \qquad \text{and} \qquad \boldsymbol{A} = \boldsymbol{CR} = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{CF} \end{bmatrix} \boldsymbol{P} \tag{4.1}$$

Again, the matrix F tells how to construct the $n-r$ dependent columns of A from the r independent columns in C. This interpretation is often missing from explanations of **rref**—it comes naturally when the process is expressed as a factorization. *C gives the independent columns of A and CF gives the dependent columns.* P orders those n columns correctly in A.

Example 2 A is a 3 by 3 matrix of rank 2. Column 3 is $-$ column $1+2$(column 2), so F contains -1 and 2. This example has $P = I$, and the zero row of R_0 is gone:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{bmatrix} \begin{bmatrix} \mathbf{1} & 0 & -1 \\ 0 & \mathbf{1} & 2 \end{bmatrix} = \boldsymbol{CR} = \boldsymbol{C} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{F} \end{bmatrix}$$

The first two columns of A are a basis for its column space. Those columns are in C. **The two rows of $\boldsymbol{R}$ are a basis for the row space of $\boldsymbol{A}$.** Those rows are independent (they contain I). They span the row space because $A = CR$ expresses every row of A as a combination of the rows of R.

Piziak and Odell [9] include $A = CR$ as the first of their "full rank factorizations" : independent columns times independent rows. Then Gram-Schmidt ($A = QR$) achieves *orthogonal columns*, and the SVD also achieves *orthogonal rows*. Those great factorizations offer a numerical stability that elimination can never match.

Here are the steps to establish $A = CR$. We know that an invertible elimination matrix E (a product of simple steps) gives $EA = R_0 = \mathbf{rref}(A)$. Then $A = E^{-1}R_0 = (m{\times}m)(m{\times}n)$. Drop the $m-r$ zero rows of R_0 and the last $m-r$ columns of E^{-1}. This leaves $A = C\begin{bmatrix} I & F \end{bmatrix} P$, where the identity matrix in R allows us to identify C in the columns of E^{-1}.

The factorization $A = CR$ reveals the first great theorem of linear algebra.

The column rank $\boldsymbol{r}$ equals the row rank.

Proof The r rows of R are independent (*from its submatrix I*). And all rows of A are combinations of the rows of R (*because* $A = CR$). Note that the rows of R belong to the row space of A because $R = (C^{\mathrm{T}}C)^{-1}C^{\mathrm{T}}A$. So the row rank is r.

The $A = CR$ factorization reveals all these essential facts and more. The equation $A\boldsymbol{x} = \boldsymbol{0}$ becomes easy to solve. *Each dependent column of A is a combination of the r independent columns in C.* That gives $n-r$ "special solutions" to $A\boldsymbol{x} = \boldsymbol{0}$: a basis for the nullspace of A.

Put those solutions into the nullspace matrix N. Recalling that $PP^{\mathrm{T}} = I$, here is $\boldsymbol{AN} =$ **zero matrix** (m by $n-r$) :

$$A = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{CF} \end{bmatrix} P \quad \text{times} \quad N = P^{\mathrm{T}} \begin{bmatrix} \boldsymbol{-F} \\ \boldsymbol{I_{n-r}} \end{bmatrix} \quad \text{equals} \quad \boldsymbol{-CF + CF = 0} \tag{4.2}$$

In Example 2 with $P = I$, N has one column $\boldsymbol{x} = \begin{bmatrix} -F & 1 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} 1 & -2 & 1 \end{bmatrix}^{\mathrm{T}}$.

The columns of N are "special" because of the submatrix I_{n-r} : One dependent column of A is a combination (given by F) of the independent columns in C.

This analysis of the $A = CR$ factorization was new to us. In a linear algebra course, it allows early examples of the column space and the crucial ideas of *linear independence* and *basis*. Those ideas appear first in examples like the matrix A above (before the student has a system for computing C). That system comes later—it is the elimination that produces $R_0 = \mathbf{rref}(A)$ and then R, with their submatrix I that identifies the columns that go into C.

The goal is to start a linear algebra class with simple examples of independence and rank and matrix multiplication. Integer matrices A reveal the *idea* of independence before the definition. And the matrix $\mathbf{rref}(A)$ now has a meaning! The identity matrix locates the first r independent columns (a basis for the column space of A). And F (r by $n-r$) gives the combinations of those r columns in C that produce the $n-r$ dependent columns CF of A.

5. The Magic Factorization $\boldsymbol{A = CW^{-1}B}$. The factorization $A = CR$ has useful properties, but symmetric treatment of columns and rows is not one of them. The rows of R were constructed by elimination. They identify the independent columns in C. There is a closely related factorization that takes r *independent rows (as well as r independent columns) directly from A*. Now an r by r submatrix $\boldsymbol{W}$ appears both in the column matrix C and the row matrix B : W is the "intersection" of r rows with r columns. So a factor W^{-1} must go in between C and B :

$$\boxed{\boldsymbol{A = CR = CW^{-1}B} \text{ as in } A = \begin{bmatrix} \mathbf{1} & \mathbf{2} & 3 \\ \mathbf{4} & \mathbf{5} & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{2} \\ \mathbf{4} & \mathbf{5} \\ 7 & 8 \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{2} \\ \mathbf{4} & \mathbf{5} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1} & \mathbf{2} & 3 \\ \mathbf{4} & \mathbf{5} & 6 \end{bmatrix}}$$

The first r independent columns and independent rows of A meet in W. A key point is that W is invertible. By interposing W^{-1} between C and B, the factorization $A{=}CW^{-1}B$ succeeds with R replaced by $W^{-1}B$.

$$\text{(5.1)} \qquad \text{Notice that} \quad \boldsymbol{WR} = \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \boldsymbol{B}.$$

THEOREM 5.1. *Suppose C contains the first r independent columns of A, with $r = \text{rank}(A)$. Suppose R contains the r nonzero rows of* $\boldsymbol{rref}(A)$. *Then* $\boldsymbol{A} = \boldsymbol{CR}$.

Now suppose the matrix B contains the first r independent rows of A. Then W is the r by r matrix where C meets B. If we look only at those r rows of $A = CR$, we see $\boldsymbol{B} = \boldsymbol{WR}$. *Since B and R both have rank r, the square matrix W must have rank r and it is invertible.*

In case A itself is square and invertible, we have $A = C = B = W$. Then the factorization $A = CW^{-1}B$ reduces to $A = WW^{-1}W$. For a rectangular matrix with independent columns and rows coming first in W, we see columns times W^{-1} times rows:

$$\text{(5.2)} \qquad A = \begin{bmatrix} W & H \\ J & K \end{bmatrix} = \begin{bmatrix} W \\ J \end{bmatrix} W^{-1} \begin{bmatrix} W & H \end{bmatrix} = CW^{-1}B$$

All this is the work of a "coordinated" person. If we ask an algebraist (as we did), the factors B and W^{-1} and C become coordinate-free linear maps that reproduce A:

$$\text{(5.3)} \qquad A : \mathbf{R}^n \to \mathbf{R}^n/\text{kernel}(A) \to \text{image}(A) \to \mathbf{R}^m$$

The map in the middle is an isomorphism between r-dimensional spaces. But the natural bases are different for those two spaces. The columns of C give the right basis for image(A) = "column space". The rows of B are natural for the dual space (row space) $\mathbf{R}^n$/kernel(A). So there is an invertible r by r matrix—a change of basis—when we introduce coordinates. *That step is represented by* W^{-1}. The invertibility of W follows from the algebra.

When we verify equation (5.3) in matrix language (with coordinates), the invertibility of the column-row intersection W will be explicitly proved. Here is that additional proof.

$\boldsymbol{A} =$ any matrix of rank r	$m \times n$	
$\boldsymbol{C} = r$ independent columns of A	$m \times r$	
$\boldsymbol{B} = r$ independent rows of A	$r \times n$	
$\boldsymbol{W} =$ intersection of C and B	$r \times r$	

$$\text{Let } \boldsymbol{A} = \begin{array}{cc} \begin{bmatrix} \boldsymbol{W} & \boldsymbol{H} \\ \boldsymbol{J} & \boldsymbol{K} \end{bmatrix} & \begin{array}{l} r \\ m-r \end{array} \\ \begin{array}{cc} r & n-r \end{array} & \end{array}$$

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{W} \\ \boldsymbol{J} \end{bmatrix} \qquad \boldsymbol{B} = \begin{bmatrix} \boldsymbol{W} & \boldsymbol{H} \end{bmatrix}$$

THEOREM 5.2. *The* $\boldsymbol{r}$ *by* $\boldsymbol{r}$ *matrix* $\boldsymbol{W}$ *also has rank* $\boldsymbol{r}$. *Then* $\boldsymbol{A = CW^{-1}B}$.

1. Combinations V of the rows of B must produce the dependent rows in $\begin{bmatrix} \boldsymbol{J} & \boldsymbol{K} \end{bmatrix}$

Then $\begin{bmatrix} \boldsymbol{J} & \boldsymbol{K} \end{bmatrix} = \boldsymbol{VB} = \begin{bmatrix} \boldsymbol{VW} & \boldsymbol{VH} \end{bmatrix}$ for some matrix V and $\boldsymbol{C} = \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{V} \end{bmatrix} \boldsymbol{W}$

2. Combinations T of the columns of C must produce the dependent columns in $\begin{bmatrix} \boldsymbol{H} \\ \boldsymbol{K} \end{bmatrix}$

Then $\begin{bmatrix} \boldsymbol{H} \\ \boldsymbol{K} \end{bmatrix} = \boldsymbol{CT} = \begin{bmatrix} \boldsymbol{WT} \\ \boldsymbol{JT} \end{bmatrix}$ for some matrix T and $\boldsymbol{B} = \boldsymbol{W} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{T} \end{bmatrix}$

3. $\boldsymbol{A} = \begin{bmatrix} W & H \\ VW & VH \end{bmatrix} = \begin{bmatrix} W & WT \\ VW & VWT \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{V} \end{bmatrix} \begin{bmatrix} \boldsymbol{W} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{T} \end{bmatrix} = \boldsymbol{CW^{-1}B}$

Each step $\mathbf{1}, \mathbf{2}, \mathbf{3}$ proves again that W is invertible. If its rank were less than r, then W could not be a factor of C or B or A. In case C and B are not in the first r columns and rows of A, permutations P_R and P_C will give $P_R\,A\,P_C = \begin{bmatrix} W & H \\ J & K \end{bmatrix}$ and the proof goes through.

Uniqueness The columns of C are a basis for the column space of A. Then the columns of R contain the unique coefficients that express each column of A in that basis: $A = CR$. But W and B are not unique. They can come from any r independent rows of A.

This is consistent with (and proves!) the uniqueness of the reduced row echelon form R_0, and the non-uniqueness of the steps from A to R_0.

$A = CR$ and $A = CW^{-1}B$ both lead to explicit expressions for the *pseudoinverse* A^+. Those factors have full rank, so we know that $A^+ = R^+C^+$ and $A^+ = B^+WC^+$. These factors are one-sided inverses as in $R^+ = R^{\mathrm{T}}(RR^{\mathrm{T}})^{-1}$ and $C^+ = (C^{\mathrm{T}}C)^{-1}C^{\mathrm{T}}$.

We don't know the full history of this column-row factorization of A into $CW^{-1}B$. But an excellent paper by Hamm and Huang [3] extends the theory to the general case when W^{-1} becomes a pseudoinverse. They analyze CUR approximations for large matrices (see below), and they have also provided a valuable set of references—including the 1956 paper [8] by Roger Penrose. That paper followed his 1955 introduction of the Moore-Penrose pseudoinverse.

6. Applications to Large Matrices. Our starting point for $A = CR$ has been its connection to the reduced row echelon form $R_0 = \mathbf{rref}(A)$. That is traditionally a classroom construction, and classrooms are usually limited to small matrices. But the idea of using actual columns and rows of the matrix A can be highly attractive. Those vectors have meaning. They are often *sparse and/or nonnegative*. They reflect useful properties that we wish to preserve in approximating A.

A persuasive and widely read essay by Lee and Seung [5] made this point strongly. And there has been good progress on the key decision: *Which k columns and k rows to choose*? It is understood that for large matrices, k may be much less than the rank r of A.

An early start was the analysis of "pseudoskeleton approximations" to an exact skeleton $A = CW^{-1}B$. This can be accurate (provided there exists a rank k matrix close to A). The optimal choice of columns and rows for C and B is connected in [2] to finding a subdeterminant of maximum volume.

Interpolative decomposition has become part of *randomized* numerical linear algebra.
Exact factors C, W, B would be submatrices of A and require no new storage. But if A itself is very large, its size will have to be reduced. The computation often samples the columns of A by $Y = AG$, for an n by k Gaussian random matrix G. Then a crucial step is to produce an orthonormal basis Q for that k-dimensional approximate column space (at reasonable cost). This is Gram-Schmidt with column pivoting, aiming to put the most important columns first: $AP = QR$.

The whole randomized algorithm is beautifully presented by Martinsson in [6] and in the extended survey [7] by Martinsson and Tropp. When the singular values of A decay rapidly, these algorithms succeed with a moderate choice of k—the computational rank that replaces the actual rank of A.

7. Implementation. An `rref` function has always been part of MATLAB. It was intended for academic use in computer experiments supplementing a traditional linear algebra course, and has been largely ignored by most users for many years. We were pleasantly surprised when the function proved to be exactly what was needed to compute the $\boldsymbol{A = CR}$ and ultimately the $\boldsymbol{A = C\,W^{-1}B}$ factorizations.

7.1. `rref`. This description of `rref` was probably written in the 1980's. It is still valid today.

```
R = rref(A) produces the reduced row echelon form of A.

[R,jb] = rref(A) also returns a vector, jb, so that:
    r = length(jb) is this algorithm's idea of the rank of A,
    x(jb) are the bound variables in a linear system, Ax = b,
    A(:,jb) is a basis for the range of A,
    R(1:r,jb) is the r-by-r identity matrix.
```

An important output is the indices `jb` of the leading `r` independent columns of `A`. The statement

```
[~,jb] = rref(A)
```

ignores the echelon form and just retains the column indices.

7.2. `cr` and `cab`. We have developed two new functions, `cr` and `cab`. The function `cr` produces two matrices. The first is `C`, a basis for the column space. The second output is `R`, the `rref` of `A` with any zero rows removed.

```
[C,R] = cr(A)
```

`rref` is used once by `cr` to find both `C` and `R`. Here is the code for `cr`.

```
[R,jb] = rref(A);
r = length(jb);       % r = rank.
R = R(1:r,:);        % R(:,jb) == eye(r).
```

```
C = A(:,jb)
```

The function `cab` produces three matrices, `C`, `W` and `B`, with

```
[C,W,B] = cab(A)  or  [C,W,B,cols,rows] = cab(A)
```

`C = A(:,cols)` is a subset of the columns and forms a basis for the column space. The same `C` is produced by `cr`. `B = A(rows,:)` is a subset of the rows and forms a basis for the row space. And `W = A(rows,cols)` is the set intersection of `C` and `B`. The original `A` can be reconstructed with

```
A = C*inv(W)*B
```

Moreover the rank of `A` is `r = rank(A) = length(cols) = length(rows)`.

Consequently `W` is square and its size is `r` -by- `r`

`rref` is used twice by `cab`. The first use finds a basis for the column space. The second use is with the transpose of A and finds a basis for the row space. The echelon forms themselves are discarded; only the pivot indices are retained. Here is the code for `cab`.

```
[~,cols] = rref(A);    % Column space
C = A(:,cols);
[~,rows] = rref(A');   % Row space
B = A(rows,:);
W = A(rows,cols);      % Intersection
```

The theorem that the column space and the row space have the same dimension implies that `cols` and `rows` have the same length and ultimately that `W` is square and nonsingular.
Its size is the rank.

7.3. Gauss-Jordan. The algorithm used by `rref` is known in numerical analysis as Gauss-Jordan elimination. When Gauss-Jordan is applied to an augmented matrix `[A b]`, it solves the linear system `A*x = b`. The Gaussian elimination algorithm is less work. It does not zero elements above the pivot. It stops with an upper triangular factor `U` and then solves the modifed augmented system by back substitution.

Most modern computers have a fused multiply-add instruction, `FMA`, that multiplies two floating point values and then adds a third value to the result in one operation. For an n-by-n system, Gauss-Jordan requires about `(1/2)*n^3` FMAs, while Gaussian elimination requires less, only `(1/3)*n^3` FMAs. This is one of the reasons why the numerical linear algebra community has been less interested in `rref`.

7.4. Examples. Magic squares provide good examples of our factorizations. An n-by-n magic square is a matrix whose elements are the integers from 1 to n^2 and whose row sums, column sums, and sums along both principal diagonals are all the same magic sum. Many magic squares are rank deficient.

A 4-by-4 magic square is one of the mathematical objects in Melencolia I, a 1514 engraving by the German Renaissance artist Albrecht Dürer.

```
A =
    16     3     2    13
     5    10    11     8
     9     6     7    12
     4    15    14     1
```

This matrix is rank deficient, but the linear dependencies are not obvious. The rank is revealed to be three by the `cr` function.

```
[C,R] = cr(A)

C =

    16     3     2
     5    10    11
     9     6     7
     4    15    14

R =

     1     0     0     1
     0     1     0    -3
     0     0     1     3
```

The matrix `R` is the `rref` of `A` with a zero row removed.

Equal treatment of rows and columns is provided by the `cab` function.

```
[C,W,B] = cab(A) = C*inv(W)*B.
```

```
C =
    16     3     2
     5    10    11
     9     6     7
     4    15    14
```

```
W =
    16     3     2
     5    10    11
     9     6     7

B =
    16     3     2    13
     5    10    11     8
     9     6     7    12
```

The Franklin *semimagic* square is attributed to Benjamin Franklin in c.1752 [1]. Its rows and columns all have the required magic sum, but the diagonals do not, so it isn't fully magic. However, many other interesting submatrices are magic, including bent diagonals and any eight elements arranged symmetrically about the center.

```
A =
    52    61     4    13    20    29    36    45
    14     3    62    51    46    35    30    19
    53    60     5    12    21    28    37    44
    11     6    59    54    43    38    27    22
    55    58     7    10    23    26    39    42
     9     8    57    56    41    40    25    24
    50    63     2    15    18    31    34    47
    16     1    64    49    48    33    32    17
```

It is hard to think of Franklin's square as a linear transformation, but `cab` can still compute its rank.

```
[C,W,B] = cab(A)

C =
    52    61     4
    14     3    62
    53    60     5
    11     6    59
    55    58     7
     9     8    57
    50    63     2
    16     1    64

W =
    52    61     4
    14     3    62
    53    60     5

B =
    52    61     4    13    20    29    36    45
    14     3    62    51    46    35    30    19
    53    60     5    12    21    28    37    44
```

So the rank is three. And, `C*inv(W)*B` is within roundoff error of `A`.

```
test = C*inv(W)*B
```

```
test =
     52.00   61.00    4.00   13.00   20.00  29.00  36.00  45.00
     14.00    3.00   62.00   51.00   46.00  35.00  30.00  19.00
     53.00   60.00    5.00   12.00   21.00  28.00  37.00  44.00
     11.00    6.00   59.00   54.00   43.00  38.00  27.00  22.00
     55.00   58.00    7.00   10.00   23.00  26.00  39.00  42.00
      9.00    8.00   57.00   56.00   41.00  40.00  25.00  24.00
     50.00   63.00    2.00   15.00   18.00  31.00  34.00  47.00
     16.00    1.00   64.00   49.00   48.00  33.00  32.00  17.00
```

Rounding `test` to the nearest integers reproduces Franklin's magic square exactly.

REFERENCES

[1] Benjamin Franklin, *The Papers of Benjamin Franklin* **4** : 1750-1753, To Peter Collinson (American Philosophical Society and Yale University) 392. **https://franklinpapers.org/framedVolumes.jsp?vol=4&page=392a**. Also printed in *Experiments and Observations on Electricity* (London, 1769), 350-354.

[2] S. A. Goreinov, E. E. Tyrtyshnikov, and N. L. Zamarashkin, *A theory of pseudoskeleton approximation,* Linear Algebra and Its Applications **261** (1997) 1-21.

[3] K. Hamm and L. Huang, *Perspectives on CUR decompositions,* Applied and Comp. Harmonic Analysis **48** (2020) 1088-1099 and arXiv : 1907.12668.

[4] R. Kannan and S. Vempala, *Randomized algorithms in numerical linear algebra,* Acta Numerica **26** (2017) 95-135.

[5] D. D. Lee and H. S. Seung, *Learning the parts of objects by nonnegative matrix factorization,* Nature **401** (1999) 788-791.

[6] P. G. Martinsson, *Randomized methods for matrix computations,* in *The Mathematics of Data,* American Mathematical Society (2018).

[7] Per-Gunnar Martinsson and Joel Tropp, *Randomized numerical linear algebra : Foundations and Algorithms,* Acta Numerica (2020) and arXiv : 2002.01387.

[8] R. Penrose, *On best approximate solutions of linear matrix equations,* Math. Proc. Cambridge Phil. Soc. **52** (1956) 17-19.

[9] R. Piziak and P. L. Odell, *Full rank factorization of matrices,* Mathematics Magazine **72** (1999) 193-201.

[10] D. C. Sorensen and Mark Embree, *A DEIM-induced CUR factorization,* SIAM J. Scientific Computing **38** (2016) 1454-1482 and arXiv : 1407.5516.

[11] Gilbert Strang, *Linear Algebra for Everyone,* Wellesley-Cambridge Press (2020).

[12] Gilbert Strang and Steven Lee, *Row reduction of a matrix and $A = CaB$,* American Mathematical Monthly **107** (2000) 681-688.

[13] S. Voronin and P. G. Martinsson, *Efficient algorithms for CUR and interpolative matrix decompositions,* Advances in Computational Mathematics **43** (2017) 495-516.

Acknowledgments. We thank David Vogan for translating $CW^{-1}B$ into algebra in equation (5.3), and Dan Drucker, Steven Lee, and Alexander Lin for a stream of ideas and references about CR and $CW^{-1}B$. A joint paper on their construction and applications is in preparation.