Elimination and Factorization $A = CR$

Gilbert Strang

Abstract If a matrix A has rank r, then its row echelon form (from elimination) contains the identity matrix in its first r independent columns. How do we *interpret the matrix* F that appears in the remaining columns of that echelon form? F multiplies those first r independent columns of A to give its $n-r$ dependent columns. Then F reveals bases for the row space and the nullspace of the original matrix A. And F is the key to the column-row factorization $\boldsymbol{A = CR}$.

1. Elimination must be just about the oldest algorithm in linear algebra. By systematically producing zeros in a matrix, it simplifies the solution of m equations $A\boldsymbol{x} = \boldsymbol{b}$. We take as example this 3 by 4 matrix A, with row 1 + row 2 = row 3. Then its rank is $r = 2$, and we execute the familiar elimination steps to find its *reduced row echelon form* Z:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 11 & 17 \\ 3 & 7 & 37 & 57 \\ 4 & 9 & 48 & 74 \end{bmatrix} \rightarrow \boldsymbol{Z} = \begin{bmatrix} 1 & 0 & 3 & 5 \\ 0 & 1 & 4 & 6 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

At this point, we pause the algorithm to ask a question: **How is $\boldsymbol{Z}$ related to $\boldsymbol{A}$?** One answer comes from the fundamental subspaces associated with A:

1) The two nonzero rows of $\boldsymbol{Z}$ (call them $\boldsymbol{R}$) are a basis for the row space of A.

2) The first two columns of $\boldsymbol{A}$ (call them $\boldsymbol{C}$) are a basis for the column space of A.

3) The nullspace of $\boldsymbol{Z}$ equals the nullspace of $\boldsymbol{A}$ (orthogonal to the same row space).

Those were our reasons for elimination in the first place. "Simplify the matrix A without losing the information it contains." By applying the same steps to the right hand side of $A\boldsymbol{x} = \boldsymbol{b}$, we reach an equation $Z\boldsymbol{x} = \boldsymbol{d}$—with the same solutions $\boldsymbol{x}$ and the simpler matrix Z.

The object of this short note is to express the result of elimination in a different way. This factorization cannot be new, but it deserves new emphasis.

Elimination factors $\boldsymbol{A}$ into $\boldsymbol{C}$ times $\boldsymbol{R} = (\boldsymbol{m} \times \boldsymbol{r})$ times $(\boldsymbol{r} \times \boldsymbol{n})$

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 11 & 17 \\ 3 & 7 & 37 & 57 \\ 4 & 9 & 48 & 74 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 7 \\ 4 & 9 \end{bmatrix} \begin{bmatrix} 1 & 0 & 3 & 5 \\ 0 & 1 & 4 & 6 \end{bmatrix} = \boldsymbol{CR}$$

C has full column rank $r = 2$ and R has full row rank $r = 2$. When we establish that $A = CR$ is true for every matrix A, this factorization brings with it a proof of the first great theorem in linear algebra: **Column rank equals row rank**.

2. Here is a description of C and R that is independent of the algorithm (row operations) that computes them.

Suppose the first r independent columns of A go into C. Then the other $n - r$ columns of A must be combinations CF of those independent columns in C. That key matrix F is part of the row factor $R = \begin{bmatrix} I & F \end{bmatrix} P$, with r independent rows. Then right away we have $A = CR$:

$$\boldsymbol{A = CR = \begin{bmatrix} C & CF \end{bmatrix} P = \begin{bmatrix} \text{Independent cols} & \text{Dependent cols} \end{bmatrix} \text{Permute cols}} \tag{1}$$

If the r independent columns come first in A, that permutation matrix will be $P = I$. Otherwise we need P to permute the n columns of C and CF into correct position in A.

Here is an example of $A = \boldsymbol{C}\begin{bmatrix} I & F \end{bmatrix} P$ in which P exchanges columns 2 and 3:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} P = \boldsymbol{CR}. \tag{2}$$

The essential information in $Z = \mathbf{rref}(A)$ is the list of r independent columns of A, and the matrix F (r by $n - r$) that combines those independent columns to give the $n - r$ dependent columns CF in A. This uniquely defines Z in equation (3).

3. The factorization $A = CR$ is confirmed. But how do we determine the first r independent columns in A (going into C) and the dependencies CF of the remaining $n - r$ columns? This is the moment for **row operations** on A. Three operations are allowed, to put A into its reduced row echelon form $Z = \mathbf{rref}(A)$:

(a) Subtract a multiple of one row from another row (below or above)

(b) Exchange two rows

(c) Divide a row by its first nonzero entry

All linear algebra teachers and a positive fraction of students know those steps and their outcome $\mathbf{rref}(A)$. It contains an r by r identity matrix I (only zeros can precede those 1's). The position of I reveals the first r independent columns of A. And equation (1) above reveals the meaning of F! It tells us how the $n-r$ *dependent* columns CF of A come from the independent columns in C. The remaining $m - r$ dependent rows of A must become zero rows in Z:

$$\textbf{Elimination reduces } \boldsymbol{A} \textbf{ to } \boldsymbol{Z} = \mathbf{rref}(A) = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{F} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \boldsymbol{P} \tag{3}$$

All our row operations (a)(b)(c) are invertible. (This is Gauss-Jordan elimination: operating on rows above the pivot row as well as below.) But the matrix that reduces A to this echelon form is less important than the **factorization $\boldsymbol{A = CR}$** that it uncovers in equation (1).

4. Before we apply $A = CR$ to solving $A\boldsymbol{x} = \mathbf{0}$, we want to give a column by column (left to right) construction of **rref**(A) from A. After elimination on k columns, that part of the matrix is in its own **rref** form. We are ready for elimination on the current column $k+1$. This new column has an upper part $\boldsymbol{u}$ and a lower part $\boldsymbol{\ell}$:

$$\textbf{First } \boldsymbol{k+1} \textbf{ columns} \qquad \begin{bmatrix} I_k & F_k \\ 0 & 0 \end{bmatrix} P_k \text{ followed by } \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{\ell} \end{bmatrix}. \tag{4}$$

The big question is: **Does this new column $k+1$ join with I_k or F_k?**

If $\boldsymbol{\ell}$ is all zeros, the new column is **dependent** on the first k columns. Then $\boldsymbol{u}$ joins with F_k to produce F_{k+1} in the next step to column $k+2$.

If $\boldsymbol{\ell}$ is not all zero, the new column is **independent** of the first k columns. Pick any nonzero in $\boldsymbol{\ell}$ (preferably the largest) as the pivot. Move that row of A to the top of $\boldsymbol{\ell}$. Then use that pivot row to zero out (by standard elimination) all the rest of column $k+1$. (That step is expected to change the columns after $k+1$.) Column $k+1$ joins with I_k to produce I_{k+1}. We adjust P_k and we are ready for column $k+2$.

At the end of elimination, we have a most desirable list of column numbers. They tell us the **first r independent columns of A**. Those are the columns of C. They led to the identity matrix $I_{r \text{ by } r}$ in the row factor R of $A = CR$.

5. What is achieved by reducing A to **rref**(A)? The row space is not changed! Then its orthogonal complement (**the nullspace of A**) is not changed. Each column of CF tells us how a dependent column of A is a combination of the independent columns in C. Effectively, **the columns of F are telling us $n-r$ solutions to $A\boldsymbol{x} = \mathbf{0}$**. This is easiest to see by example.

$$\begin{matrix} x_1 + 2x_2 + 11x_3 + 17x_4 = 0 \\ 3x_1 + 7x_2 + 37x_3 + 57x_4 = 0 \end{matrix} \quad \text{reduces to} \quad \begin{bmatrix} I & F \end{bmatrix} \boldsymbol{x} = \begin{matrix} x_1 + 3x_3 + 5x_4 = 0 \\ x_2 + 4x_3 + 6x_4 = 0 \end{matrix}$$

The solution with $x_3 = 1$ and $x_4 = 0$ is $\boldsymbol{x} = \begin{bmatrix} -3 & -4 & 1 & 0 \end{bmatrix}^{\mathrm{T}}$. Notice 3 and 4 from F. The second solution with $x_3 = 0$ and $x_4 = 1$ is $\boldsymbol{x} = \begin{bmatrix} -5 & -6 & 0 & 1 \end{bmatrix}^{\mathrm{T}}$. Those solutions are the two columns of $\boldsymbol{X}$ in $\boldsymbol{AX} = \mathbf{0}$. (This example has $P = I$.) **The $n-r$ columns of $\boldsymbol{X}$ are a natural basis for the nullspace of $\boldsymbol{A}$:**

$$\boldsymbol{A} = \boldsymbol{C}\begin{bmatrix} \boldsymbol{I} & \boldsymbol{F} \end{bmatrix}\boldsymbol{P} \text{ multiplies } \boldsymbol{X} = \boldsymbol{P}^{\mathbf{T}}\begin{bmatrix} \boldsymbol{-F} \\ \boldsymbol{I_{n-r}} \end{bmatrix} \text{ to give } \boldsymbol{AX} = \boldsymbol{-CF + CF} = \mathbf{0}. \tag{5}$$

With $PP^{\mathrm{T}} = I$ for the permutation, each column of X solves $A\boldsymbol{x} = \mathbf{0}$. Those $n-r$ solutions in X tell us what we know: Each dependent column of A is a combination of the independent columns in C. In the example, column 3 = 3 (column 1) + 4 (column 2).

Gauss-Jordan elimination leading to $A = CR$ is less efficient than the Gauss process that directly solves $A\boldsymbol{x} = \boldsymbol{b}$. The latter stops at a triangular system $U\boldsymbol{x} = \boldsymbol{c}$: back substitution creates $\boldsymbol{x}$. Gauss-Jordan has the extra cost of eliminating upwards. If we only want to solve equations, stopping at a triangular factorization is faster.

6. Block elimination The row operations that reduce A to its echelon form produce one zero at a time. A key part of that echelon form is an r by r identity matrix. If we think on a larger scale—instead of one row at a time—**that output $\boldsymbol{I}$ tells us that some $\boldsymbol{r}$ by $\boldsymbol{r}$ matrix has been inverted**. Following that lead brings a "matrix understanding" of elimination.

Suppose that the matrix $\boldsymbol{W}$ in the first r rows and columns of A is invertible. Then *elimination takes all its instructions from* W! One entry at a time—or all at once by "block elimination"—**$\boldsymbol{W}$ will change to $\boldsymbol{I}$**. In other words, the first r rows of A will yield I and F. **This identifies $\boldsymbol{F}$ as $\boldsymbol{W^{-1}H}$**. And the last $m - r$ rows will become zero rows.

$$\textbf{Block elimination} \quad \boldsymbol{A} = \begin{bmatrix} \boldsymbol{W} & \boldsymbol{H} \\ \boldsymbol{J} & \boldsymbol{K} \end{bmatrix} \text{ reduces to } \begin{bmatrix} \boldsymbol{I} & \boldsymbol{F} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} = \textbf{rref}(\boldsymbol{A}). \quad (6)$$

That just expresses the facts of linear algebra: If A begins with r independent rows and its rank is r, then the remaining $m - r$ rows are combinations of those first r rows : $\begin{bmatrix} J & K \end{bmatrix} = JW^{-1} \begin{bmatrix} W & H \end{bmatrix}$. Those $m - r$ rows become zero rows in elimination.

In general that first r by r block might not be invertible. But elimination will find W. We can move W to the upper left corner by row and column permutations $\boldsymbol{P_r}$ and $\boldsymbol{P_c}$.Then the full expression of block elimination to reduced row echelon form is

$$\boldsymbol{P_r A P_c} = \begin{bmatrix} \boldsymbol{W} & \boldsymbol{H} \\ \boldsymbol{J} & \boldsymbol{K} \end{bmatrix} \rightarrow \begin{bmatrix} \boldsymbol{I} & \boldsymbol{W^{-1}H} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \quad (7)$$

7. This raises an interesting point. Since A has rank r, we know that it has r independent rows and r independent columns. Suppose those rows are in a submatrix B and those columns are in a submatrix C. Is it always true that the r by r "intersection" W of those rows with those columns will be **invertible**? (Everyone agrees that somewhere in A there is an r by r invertible submatrix. The question is whether $B \cap C$ can be counted on to provide such a submatrix.) Is W automatically full rank?

The answer is *yes*. **The intersection of $\boldsymbol{r}$ independent rows of $\boldsymbol{A}$ with $\boldsymbol{r}$ independent columns does produce a matrix $\boldsymbol{W}$ of rank $\boldsymbol{r}$. $\boldsymbol{W}$ is invertible.**

Proof: Every column of A is a combination of the r columns of C.
Every column of the r by n matrix B is the same combination of the r columns of W.
Since B has rank r, its column space is all of $\mathbf{R}^{\boldsymbol{r}}$.
Then the column space of W is also $\mathbf{R}^{\boldsymbol{r}}$ and **the square submatrix $\boldsymbol{W}$ has rank $\boldsymbol{r}$**.