Multiplying and Factoring Matrices

Gilbert Strang
MIT
gs@math.mit.edu

I believe that the right way to understand matrix multiplication is **columns times rows**:

$$AB = \begin{bmatrix} & & \\ \boldsymbol{a}_1 \dots \boldsymbol{a}_n \\ & & \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{b}_n^{\mathrm{T}} \end{bmatrix} = \boldsymbol{a}_1 \boldsymbol{b}_1^{\mathrm{T}} + \cdots + \boldsymbol{a}_n \boldsymbol{b}_n^{\mathrm{T}}. \tag{1}$$

Each column $\boldsymbol{a}_k$ of an m by n matrix multiplies a row of an n by p matrix. The product $\boldsymbol{a}_k \boldsymbol{b}_k^{\mathrm{T}}$ is an m by p matrix of ***rank one***. The sum of those rank one matrices is AB.

All columns of $\boldsymbol{a}_k \boldsymbol{b}_k^{\mathrm{T}}$ are multiples of $\boldsymbol{a}_k$, all rows are multiples of $\boldsymbol{b}_k^{\mathrm{T}}$. The i, j entry of this rank one matrix is $a_{ik} b_{kj}$. The sum over k produces the i, j entry of AB in "the old way." Computing with numbers, I still find AB by **rows times columns** (inner products)!

The central ideas of matrix analysis are perfectly expressed as matrix factorizations:

$$\boxed{\boldsymbol{A = LU} \qquad \boldsymbol{A = QR} \qquad \boldsymbol{S = Q \Lambda Q^{\mathrm{T}}} \qquad \boldsymbol{A = X \Lambda Y^{\mathrm{T}}} \qquad \boldsymbol{A = U \Sigma V^{\mathrm{T}}}}$$

The last three, with eigenvalues in Λ and singular values in Σ, are often seen as column-row multiplications (a sum of outer products). The spectral theorem for S is a perfect example. The first two are Gaussian elimination ($\boldsymbol{LU}$) and Gram-Schmidt orthogonalization ($\boldsymbol{QR}$). We aim to show that those are also clearly described using rank one matrices.

The spectral theorem $\quad \boldsymbol{S = Q \Lambda Q^{\mathrm{T}}}$ $\quad$ A real symmetric matrix S is diagonalized by its orthonormal eigenvectors. The eigenvalues λ_i enter the diagonal matrix Λ. They multiply the eigenvectors $\boldsymbol{q}_i$ in the columns of Q. Then $\lambda_i \boldsymbol{q}_i$ is a column of $Q\Lambda$. The column-row multiplication $(Q\Lambda)Q^{\mathrm{T}}$ has the familiar form

$$S = \lambda_1 \boldsymbol{q}_1 \boldsymbol{q}_1^{\mathrm{T}} + \cdots + \lambda_n \boldsymbol{q}_n \boldsymbol{q}_n^{\mathrm{T}}. \tag{2}$$

To see that S times $\boldsymbol{q}_j$ produces $\lambda_j \boldsymbol{q}_j$, multiply every term $\lambda_i \boldsymbol{q}_i \boldsymbol{q}_i^{\mathrm{T}}$ by $\boldsymbol{q}_j$. By orthogonality, the only surviving term has $i = j$. That term is $\lambda_j \boldsymbol{q}_j$ because $\boldsymbol{q}_j^{\mathrm{T}} \boldsymbol{q}_j = 1$.

Of course the proof of the spectral theorem requires construction of the $\boldsymbol{q}_j$.

Elimination $\quad \boldsymbol{A = LU}$ is the result of Gaussian elimination in the usual order, starting with an invertible matrix A and ending with an upper triangular U. The key idea is that the matrix L linking U to A contains the ***multipliers*** — the numbers ℓ_{ij} that multiply row j when it is subtracted from row $i > j$ to produce $U_{ij} = 0$.

The "magic" is that those separate steps do not interfere, when we undo elimination and bring U back to A. The numbers ℓ_{ij} fall into place in L—but that key fact can take patience to verify in a classroom. Here we look for a different approach. The column-times-row idea makes the steps of elimination transparently clear.

Step 1 *of elimination* Row 1 of U is row 1 of A. Column 1 of L is column 1 of A, divided by the first pivot a_{11} (so that $\ell_{11} = 1$). Then the product $\boldsymbol{\ell}_1 \boldsymbol{u}_1^{\mathrm{T}}$ extends the first row and column of A to a rank-one matrix. So the difference is a matrix A_2 of size $n-1$ bordered by zeros in row 1 and column 1 :

$$\textbf{Step 1} \qquad \boldsymbol{A} = \boldsymbol{\ell}_1 \boldsymbol{u}_1^{\mathrm{T}} + \begin{bmatrix} 0 & \mathbf{0^T} \\ \mathbf{0} & \boldsymbol{A_2} \end{bmatrix}. \tag{3}$$

Step 2 acts in the same way on A_2. The row $\boldsymbol{u}_2^{\mathrm{T}}$ and the column $\boldsymbol{\ell}_2$ will start with a single zero. The essential point is that elimination on column 1 removed the matrix $\boldsymbol{\ell}_1 \boldsymbol{u}_1^{\mathrm{T}}$ from A.

$$A = \begin{bmatrix} 2 & 3 \\ 4 & 11 \end{bmatrix} \qquad \boldsymbol{u}_1^{\mathrm{T}} = \begin{bmatrix} 2 & 3 \end{bmatrix} \qquad \boldsymbol{\ell}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \boldsymbol{\ell}_1 \boldsymbol{u}_1^{\mathrm{T}} = \begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix}$$

$$A - \boldsymbol{\ell_1} \boldsymbol{u}_1^{\mathrm{T}} = \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{5} \end{bmatrix} = \boldsymbol{\ell_2} \boldsymbol{u}_2^{\mathrm{T}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 5 \end{bmatrix}. \text{ So } LU = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 0 & 5 \end{bmatrix}.$$

When elimination reaches A_k, there are $k-1$ zeros at the start of each row and column. Those zeros in $\boldsymbol{u}_k$ and $\boldsymbol{\ell}_k$ produce an upper triangular matrix U and a lower triangular L. The diagonal entries of U are the pivots (not zero). The diagonal entries of L are all 1's.

The linear system $A\boldsymbol{x} = \boldsymbol{b}$ is reduced to two triangular systems governed by L and U :

$$\text{Solve } L\boldsymbol{c} = \boldsymbol{b} \quad \text{and solve} \quad U\boldsymbol{x} = \boldsymbol{c}. \qquad \text{Then } A\boldsymbol{x} = LU\boldsymbol{x} = L\boldsymbol{c} = \boldsymbol{b}.$$

Forward elimination leaves $U\boldsymbol{x} = \boldsymbol{c}$, and back-substitution produces $\boldsymbol{x}$. To assure nonzero pivots, this LU decomposition requires every leading square submatrix of A (from its first k rows and columns) to be invertible.

Gram-Schmidt orthogonalization $\boldsymbol{A = QR}$ The algorithm combines independent vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ to produce orthonormal vectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$. Subtract from $\boldsymbol{a}_2$ its component in the direction of $\boldsymbol{a}_1$. Normalize at each step to unit vectors $\boldsymbol{q}$:

$$\boldsymbol{q}_1 = \frac{\boldsymbol{a}_1}{||\boldsymbol{a}_1||} = \frac{\boldsymbol{a}_1}{r_{11}} \qquad \boldsymbol{q}_2 = \frac{\boldsymbol{a}_2 - (\boldsymbol{q}_1^{\mathrm{T}} \boldsymbol{a}_2)\boldsymbol{q}_1}{||\boldsymbol{a}_2 - (\boldsymbol{q}_1^{\mathrm{T}} \boldsymbol{a}_2)\boldsymbol{q}_1||} = \frac{\boldsymbol{a}_2 - r_{12}\boldsymbol{q}_1}{r_{22}}.$$

As with elimination, this is clearer when we recover the original vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ from the final $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$:

$$\boldsymbol{a}_1 = r_{11}\boldsymbol{q}_1 \qquad \boldsymbol{a}_2 = r_{12}\, \boldsymbol{q}_1 + r_{22}\, \boldsymbol{q}_2. \tag{4}$$

In this order we see why R is triangular. At each step, $\boldsymbol{q}_1$ to $\boldsymbol{q}_k$ span the same subspace as $\boldsymbol{a}_1$ to $\boldsymbol{a}_k$. We can establish the Gram-Schmidt factorization $A = QR = \boldsymbol{q}_1 \boldsymbol{r}_1^{\mathrm{T}} + \cdots + \boldsymbol{q}_n \boldsymbol{r}_n^{\mathrm{T}}$ as follows:

The first column $\boldsymbol{q}_1$ is the first column $\boldsymbol{a}_1$ divided by its length r_{11}

The first row $\boldsymbol{r}_1^{\mathrm{T}}$ contains the inner products $\boldsymbol{q}_1^{\mathrm{T}} \boldsymbol{a}_k$.

Subtracting the rank one matrix $\boldsymbol{q}_1\boldsymbol{r}_1^{\mathrm{T}}$ leaves a matrix A_2 whose columns are **all orthogonal to $\boldsymbol{q}_1$** :

$$A = \boldsymbol{q}_1\boldsymbol{r}_1^{\mathrm{T}} + \begin{bmatrix} \mathbf{0} & A_2 \end{bmatrix}. \tag{5}$$

This is the analog of equation (3) for elimination. There we had a row of zeros above A_2. Here we have columns of A_2 orthogonal to $\boldsymbol{q}_1$. In two lines, this example reaches equation (5) :

$$A = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 \end{bmatrix} = \begin{bmatrix} 6 & 2 \\ 8 & 6 \end{bmatrix} \text{ has } r_{11} = ||\boldsymbol{a}_1|| = \mathbf{10} \text{ and unit vector } \boldsymbol{q}_1 = \frac{\boldsymbol{a}_1}{10}$$

$$r_{12} = \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a}_2 = \begin{bmatrix} 0.6 & 0.8 \end{bmatrix}\begin{bmatrix} 2 \\ 6 \end{bmatrix} = \mathbf{6} \quad \text{and} \quad A = \begin{bmatrix} 0.6 \\ 0.8 \end{bmatrix}\begin{bmatrix} 10 & 6 \end{bmatrix} + \begin{bmatrix} 0 & -1.6 \\ 0 & 1.2 \end{bmatrix}$$

That last column has length $r_{22} = \mathbf{2}$ and $A = \begin{bmatrix} 6 & 2 \\ 8 & 6 \end{bmatrix} = \begin{bmatrix} 0.6 & -0.8 \\ 0.8 & 0.6 \end{bmatrix}\begin{bmatrix} \mathbf{10} & \mathbf{6} \\ 0 & \mathbf{2} \end{bmatrix}$.

This product $A = QR$ or $Q^{\mathrm{T}}A = R$ confirms that every $r_{ij} = \boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{a}_j$ (row times column). The mysterious matrix R just contains inner products of $\boldsymbol{q}$'s and $\boldsymbol{a}$'s. R is triangular because $\boldsymbol{q}_i$ does not involve $\boldsymbol{a}_j$ for $j > i$. Gram-Schmidt uses only $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_i$ to construct $\boldsymbol{q}_i$.

The next vector $\boldsymbol{q}_2$ is the first column of A_2 divided by its length. The next vector $\boldsymbol{r}_2^{\mathrm{T}}$ contains (after a first zero) the inner products of $\boldsymbol{q}_2$ with columns of A_2 :

$$A = \boldsymbol{q}_1\boldsymbol{r}_1^{\mathrm{T}} + \boldsymbol{q}_2\boldsymbol{r}_2^{\mathrm{T}} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & A_3 \end{bmatrix}. \tag{6}$$

All columns of A_3 are orthogonal to $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$.

After n steps this is $A = QR$. Only the order of the orthogonalization steps has been modified—by subtracting components (projections) from the columns of A as soon as each new $\boldsymbol{q}_k$ direction has been found.

Now come the last two factorizations of A.

Eigenvalue Decomposition $\boldsymbol{A = X\Lambda X^{-1} = X\Lambda Y^{\mathrm{T}}}$

The effect of n independent eigenvectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ is to diagonalize the matrix A. Those "right eigenvectors" are the columns of X. Column by column, we see $AX = X\Lambda$. Then $\Lambda = X^{-1}AX$ is the diagonal matrix of eigenvalues, as usual.

To keep the balance between columns and rows, recognize that the rows of X^{-1} are the "left eigenvectors" of A. This is expressed by $\boldsymbol{X^{-1}A = \Lambda X^{-1}}$. Writing $\boldsymbol{y}_1^{\mathrm{T}}, \ldots, \boldsymbol{y}_n^{\mathrm{T}}$ for the rows of X^{-1} we have $\boldsymbol{y}_j^{\mathrm{T}}A = \lambda_j\boldsymbol{y}_j^{\mathrm{T}}$. So the diagonalization $A = X\Lambda X^{-1}$ actually has the more symmetric form $A = X\Lambda Y^{\mathrm{T}}$:

Right and left eigenvectors $$\boldsymbol{A = X\Lambda Y}^{\mathrm{T}} = \lambda_1\boldsymbol{x}_1\boldsymbol{y}_1^{\mathrm{T}} + \cdots + \lambda_n\boldsymbol{x}_n\boldsymbol{y}_n^{\mathrm{T}} \tag{7}$$

Notice that these left eigenvectors $\boldsymbol{y}_i^{\mathrm{T}}$ are normalized by $Y^{\mathrm{T}}X = X^{-1}X = I$. This requires $\boldsymbol{y}_j^{\mathrm{T}}\boldsymbol{x}_j = 1$ and confirms the biorthogonality $\boldsymbol{y}_i^{\mathrm{T}}\boldsymbol{x}_j = \boldsymbol{\delta}_{ij}$ of the two sets of eigenvectors. A symmetric matrix has $\boldsymbol{y}_j = \boldsymbol{x}_j = \boldsymbol{q}_j$ and orthonormal eigenvectors. Then $S = Q\Lambda Q^{\mathrm{T}}$.

Singular Value Decomposition $\boldsymbol{A = U\Sigma V^{\mathrm{T}}}$ By comparing with the diagonalization $A = X\Lambda X^{-1} = X\Lambda Y^{\mathrm{T}}$, we see the parallels between a right-left eigenvector decomposition (for a diagonalizable matrix) and a right-left singular value decomposition $A = U\Sigma V^{\mathrm{T}}$ (for any matrix):

$$\textbf{The SVD with singular vectors} \quad \boldsymbol{A = U\Sigma V^{\mathrm{T}}} = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \cdots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^{\mathrm{T}}. \tag{8}$$

For every matrix A, the right singular vectors in $\boldsymbol{V}$ are orthonormal and the left singular vectors $\boldsymbol{u}_j = A\boldsymbol{v}_j / ||A\boldsymbol{v}_j||$ are orthonormal. Those $\boldsymbol{v}$'s and $\boldsymbol{u}$'s are eigenvectors of $A^{\mathrm{T}}A$ and AA^{T}.

$$A^{\mathrm{T}}A\boldsymbol{v}_j = \sigma_j^2 \boldsymbol{v}_j \;\text{ and }\; (AA^{\mathrm{T}})A\boldsymbol{v}_j = \sigma_j^2 A\boldsymbol{v}_j \;\text{ and }\; (AA^{\mathrm{T}})\boldsymbol{u}_j = \sigma_j^2 \boldsymbol{u}_j. \tag{9}$$

These matrices have the same nonzero eigenvalues $\sigma_1^2, \ldots \sigma_r^2$. The ranks of A and $A^{\mathrm{T}}A$ and AA^{T} are r. When the singular values are in decreasing order $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$, the most important piece of A is $\sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}}$:

$$||A|| = \sigma_1 \;\text{ and }\; ||A - \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}}|| = \sigma_2. \tag{10}$$

The rank one matrix closest to A is $\sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}}$. The difference $A - \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}}$ will have singular values $\sigma_2 \geq \sigma_3 \geq \ldots \geq \sigma_r$. At every step—not only this first step—the SVD produces the matrix $A_k = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \cdots + \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^{\mathrm{T}}$ of rank k that is **closest to the original** A:

$$(\textbf{Eckart-Young}) \qquad \boldsymbol{||A - B|| \geq ||A - A_k|| = \sigma_{k+1}} \quad \text{if } B \text{ has rank } k. \tag{11}$$

Thus the SVD produces the rank one pieces $\sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^{\mathrm{T}}$ ***in order of importance***. This is a central result in data science. We are measuring all these matrices by their spectral norms:

$$||A|| = \text{ maximum of } ||A\boldsymbol{x}|| = \text{ maximum of } \boldsymbol{u}^{\mathrm{T}}A\boldsymbol{v} \quad \text{with } ||\boldsymbol{x}|| = ||\boldsymbol{u}|| = ||\boldsymbol{v}|| = 1.$$

In Principal Component Analysis, the leading singular vectors are "principal components". In statistics, each row of A is centered by subtracting its mean value from its entries. Then $S = AA^{\mathrm{T}}$ is a sample covariance matrix. Its top eigenvector $\boldsymbol{u}_1$ represents the combination of rows of S with the greatest variance.

Note. For very large data matrices, the SVD is too expensive to compute. An approximation takes its place. That approximation often uses the inexpensive steps of elimination! Now elimination may begin with the ***largest entry of*** A, and not necessarily with a_{11}.

Geometrically, the singular values in Σ stretch the unit circle $||\boldsymbol{x}|| = 1$ into an ellipse. The factorization $U\Sigma V^{\mathrm{T}} =$ (***orthogonal***) times (***diagonal***) times (***orthogonal***) expresses any matrix (roughly speaking) as a rotation times a stretching times a rotation. This has become central to numerical linear algebra.

It may surprise the reader (as it did the author) that the columns of X in $A = X\Lambda Y^{\mathrm{T}}$ are ***right*** eigenvectors, while the columns of U in $A = U\Sigma V^{\mathrm{T}}$ are called ***left*** singular vectors. Perhaps this just confirms that mathematics is a human and fallible (and wonderful) joint enterprise of us all.

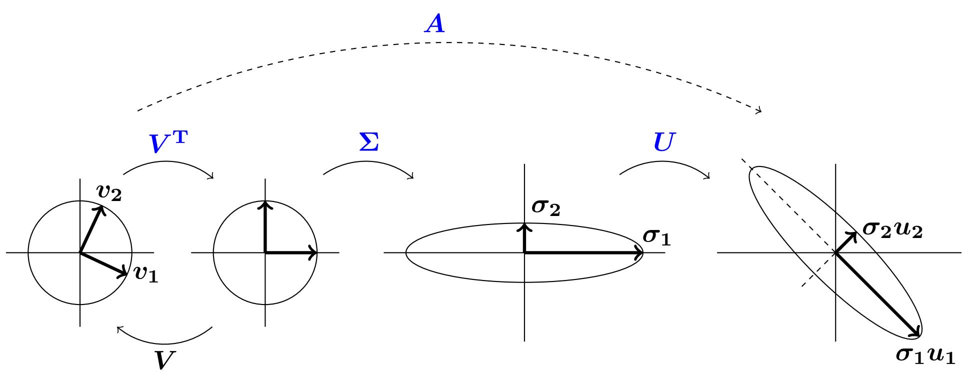

Figure 1: U and V are rotations and possible reflections. Σ stretches circle to ellipse.

Factorizations can fail ! Of the five principal factorizations, only two are guaranteed. Every symmetric matrix has the form $S = Q\Lambda Q^{\mathrm{T}}$ and every matrix has the form $A = U\Sigma V^{\mathrm{T}}$. The cases of failure are important too (or adjustment more than failure). $A = LU$ now requires an "echelon form E" and diagonalization needs a "Jordan form J". Matrix multiplication is still columns times rows.

Elimination to row reduced echelon form $\quad \boldsymbol{A} = \boldsymbol{CE} = (m \text{ by } r)(r \text{ by } n)$.

The rank of all three matrices is r. E normally comes from operations on the rows of A. Then it may have zero rows. If we work instead with columns of A, the factors C and E have direct meaning.

C contains r independent columns of A (a basis for the column space of A).

E expresses each column of A as a combination of the basic columns in C.

To choose those independent columns, work from left to right ($j = 1$ to $j = n$).

A column of A is *included* in C when it is not a combination of preceding columns.

The r corresponding columns of E contain the r by r identity matrix.

A column of A is *excluded* from C when it *is* a combination of preceding columns of C.

The corresponding column of E contains the coefficients in that combination.

Example $$\boldsymbol{A} = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix} = \boldsymbol{CE}$$

The entries of E are uniquely determined because C has independent columns.

The remarkable point is that **this E coincides with the row reduced echelon form of** A—except that zero rows are here discarded. Each "1" from the identity matrix inside E is the first nonzero in that row of E. Those 1's appear in descending order in I and E. The key to the rest of E is this :

> The nullspace of E is the nullspace of $A = CE$. (C has independent columns.)
> Therefore the row space of E is the row space of A.
> Then our E must be the row reduced echelon form without its zero rows.

The nullspace and row space of any matrix are orthogonal complements. So the column construction and the row construction must find the same E.

Gram-Schmidt echelon form $\boldsymbol{A = QU}$ $= (m$ by $r)(r$ by $n)$.

When the columns of A are linearly dependent ($r < n$), Gram-Schmidt breaks down. Only r orthonormal columns $\boldsymbol{q}_j$ are combinations of columns of A. Those $\boldsymbol{q}_j$ are combinations of the independent columns in C above, because Gram-Schmidt also works left to right. Then $A = CE$ is the same as Gram-Schmidt $C = QR$ multiplied by E :

$$A = CE = (QR)E = Q(RE) = QU. \tag{12}$$

Q has r orthonormal columns : $Q^{\mathrm{T}}Q = I$. The upper triangular $U = RE$ combines an r by r invertible triangular matrix U from Gram-Schmidt and the r by n echelon matrix E from elimination. The nullspaces of U and E and A are all the same, and $A = QU$.

The Jordan form $\boldsymbol{A = GJG^{-1} = GJH^{\mathrm{T}}}$ of a square matrix

Now it is not the columns of A but its eigenvectors that fail to span $\mathbf{R}^n$. We need to supplement the eigenvectors by "generalized eigenvectors" :

$$A\boldsymbol{g}_j = \lambda_j \boldsymbol{g}_j \quad \text{is supplemented as needed by} \quad A\boldsymbol{g}_k = \lambda_k \boldsymbol{g}_k + \boldsymbol{g}_{k-1}$$

The former puts λ_j on the diagonal of J. The latter produces also a "1" on the superdiagonal. The construction of the Jordan form J is an elegant mess (and I believe that a beginning linear algebra class has more important things to do : the five factorizations).

The only novelty is to see ***left generalized eigenvectors*** $\boldsymbol{h}_i^{\mathrm{T}}$ when we invert G. Start from a 2 by 2 Jordan block :

$$\boldsymbol{AG} = A\begin{bmatrix} \boldsymbol{g}_1 & \boldsymbol{g}_2 \end{bmatrix} = \begin{bmatrix} \lambda\boldsymbol{g}_1 & \lambda\boldsymbol{g}_2 + \boldsymbol{g}_1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{g}_1 & \boldsymbol{g}_2 \end{bmatrix}\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} = \boldsymbol{GJ}. \tag{13}$$

Then $AG = GJ$ gives $G^{-1}A = JG^{-1}$. Write $\boldsymbol{h}^{\mathrm{T}}$ for the rows of G^{-1} (the left generalized eigenvectors) :

$$\begin{bmatrix} \boldsymbol{h}_1^{\mathrm{T}} \\ \boldsymbol{h}_2^{\mathrm{T}} \end{bmatrix} A = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}\begin{bmatrix} \boldsymbol{h}_1^{\mathrm{T}} \\ \boldsymbol{h}_2^{\mathrm{T}} \end{bmatrix} \quad \text{is} \quad \boldsymbol{h}_1^{\mathrm{T}}A = \lambda\boldsymbol{h}_1^{\mathrm{T}} + \boldsymbol{h}_2^{\mathrm{T}} \quad \text{and} \quad \boldsymbol{h}_2^{\mathrm{T}}A = \lambda\boldsymbol{h}_2^{\mathrm{T}}. \tag{14}$$

In the same way that $A = X\Lambda X^{-1}$ became $A = X\Lambda Y^{\mathrm{T}}$, the Jordan decomposition $A = GJG^{-1}$ has become $A = GJH^{\mathrm{T}}$. We have rows times columns :

$$A = \lambda\boldsymbol{g}_1\boldsymbol{h}_1^{\mathrm{T}} + (\lambda\boldsymbol{g}_2 + \boldsymbol{g}_1)\boldsymbol{h}_2^{\mathrm{T}} = \boldsymbol{g}_1(\lambda\boldsymbol{h}_1^{\mathrm{T}} + \boldsymbol{h}_2^{\mathrm{T}}) + \lambda\boldsymbol{g}_2\boldsymbol{h}_2^{\mathrm{T}}. \tag{15}$$

Summary The first lines of this paper connect inner and outer products. Those are "rows times columns" and "columns times rows": Level 1 multiplication and Level 3 multiplication.

Level 1 **Inner product $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}$:** row times column $= \begin{bmatrix} a_1 & \dots & a_n \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} =$ **scalar**

Level 2 **Linear combination $A\boldsymbol{b}$:** $\begin{bmatrix} \boldsymbol{a}_1 & \dots & \boldsymbol{a}_n \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \sum \boldsymbol{a}_j b_j \quad =$ **vector**

Level 3 **Outer product $\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}$:** column times row $\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \begin{bmatrix} b_1 & \dots & b_n \end{bmatrix} \quad =$ **matrix**

The product AB of n by n matrices can be computed at every level, always with the same n^3 multiplications:

Level 1 (n^2 inner products) (n multiplications each) $= \boldsymbol{n^3}$
Level 2 (n columns $A\boldsymbol{b}$) (n^2 multiplications each) $= \boldsymbol{n^3}$
Level 3 (n outer products) (n^2 multiplications each) $= \boldsymbol{n^3}$

These correspond to the three levels of Basic Linear Algebra Subroutines (BLAS). Those are the core operations in LAPACK at the center of computational linear algebra [1]. Our factorizations are produced by four of the most frequently used MATLAB commands: **lu, gr, eig,** and **svd**

Finally we verify the most important property of matrix multiplication.

The associative law is $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$

Multiplying columns times rows satisfies this fundamental law. The matrices A, B, C are m by n, n by p, and p by q. When $n = p = 1$ and B is a scalar b_{jk}, the laws of arithmetic give two equal matrices of rank one:

$$(\boldsymbol{a}_j\, b_{jk})\, \boldsymbol{c}_k^{\mathrm{T}} = \boldsymbol{a}_j (b_{jk}\, \boldsymbol{c}_k^{\mathrm{T}}). \tag{16}$$

The full law $(AB)C = A(BC)$ will follow from the agreement of double sums:

$$\sum_{k=1}^{p} \sum_{j=1}^{n} (\boldsymbol{a}_j\, b_{jk})\, \boldsymbol{c}_k^{\mathrm{T}} = \sum_{j=1}^{n} \sum_{k=1}^{p} \boldsymbol{a}_j (b_{jk}\, \boldsymbol{c}_k^{\mathrm{T}}). \tag{17}$$

We are just adding the same np terms. After the inner sum on each side, this becomes

$$\sum_{k=1}^{p} (AB)_k\, \boldsymbol{c}_k^{\mathrm{T}} = \sum_{j=1}^{n} \boldsymbol{a}_j\, (BC)_j^{\mathrm{T}}. \tag{18}$$

With another column-row multiplication this is $(AB)C = A(BC)$. Parentheses are not needed in $Q\Lambda Q^{\mathrm{T}}$ and $X\Lambda Y^{\mathrm{T}}$ and $U\Sigma V^{\mathrm{T}}$.

REFERENCES

1. E. Anderson et al., LAPACK Users' Guide, 3rd edition, SIAM, 1999.
2. G. Strang, *Introduction to Linear Algebra*, 5th edition, Wellesley-Cambridge Press, 2016.
3. L. N. Trefethen and D. Bau, *Numerical Linear Algebra*, SIAM, 1997.