New Ideas in "Linear Algebra for Everyone"

Gilbert Strang

These notes are a chapter-by-chapter comparison of the 2020 *Linear Algebra for Everyone* with the 2016 *Introduction to Linear Algebra, 5th edition.* Both are full textbooks for a linear algebra course. Both books include important applications to least squares and differential equations. Eigenvalues lead directly to singular values.

Chapter 1 The course begins with vectors. Their combinations $c\boldsymbol{v} + d\boldsymbol{w}$ fill a plane. Their dot products give length and angle: $\boldsymbol{v} \cdot \boldsymbol{w} = ||\boldsymbol{v}||\,||\boldsymbol{w}|| \cos\theta$. Then matrices multiply vectors in two ways: $A\boldsymbol{x}$ contains dot products of $\boldsymbol{x}$ with the **rows** of A, and also $A\boldsymbol{x}$ is a **combination of the columns of $\boldsymbol{A}$**. The first way is for hand computation. The second way is for understanding.

LAFE extends those linear combinations $A\boldsymbol{x}$ to matrix multiplication AB in Chapter 1:

Column of $\boldsymbol{AB} = \boldsymbol{A}$ times column of $\boldsymbol{B}$ = combination of the columns of $\boldsymbol{A}$.

Those combinations of columns lead directly to essential ideas. **This is our new start**.

1 Independent columns versus dependent columns

2 The number of independent columns (**the rank of $\boldsymbol{A}$**)

3 All combinations of the columns (**the column space of $\boldsymbol{A}$**)

4 $\boldsymbol{A = CR}$ All columns of $\boldsymbol{A}$ from the independent columns in $\boldsymbol{C}$

$$\boldsymbol{A} = \begin{bmatrix} \mathbf{1} & \mathbf{3} & \mathbf{4} \\ \mathbf{2} & \mathbf{2} & \mathbf{4} \\ \mathbf{5} & \mathbf{1} & \mathbf{6} \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 2 & 2 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{1} \\ \mathbf{0} & \mathbf{1} & \mathbf{1} \end{bmatrix} = \boldsymbol{CR} \qquad \begin{matrix} \text{Column 3} = \text{Columns 1 + 2} \\ \text{Column space} = \text{plane in } \mathbf{R}^3 \\ A, C, R \text{ all have rank 2} \end{matrix}$$

The examples are small matrices of integers. All students detect dependent columns. The special case of rank 1 has one independent column in C. Then R has one row. The great fact that column rank = row rank becomes clear for rank 1:

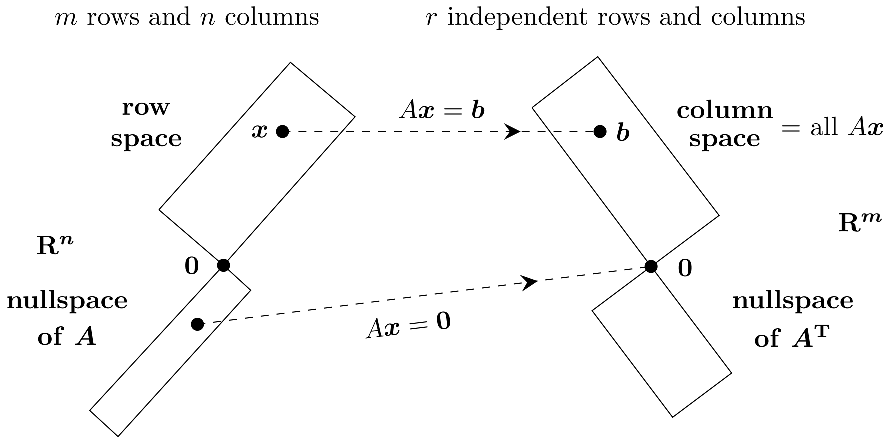
$$\boldsymbol{A} = \begin{bmatrix} \mathbf{2} & \mathbf{4} & \mathbf{10} \\ \mathbf{3} & \mathbf{6} & \mathbf{15} \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{2} & \mathbf{5} \end{bmatrix} = \boldsymbol{CR} \qquad \begin{matrix} \textbf{Columns in the same direction} \\ \textbf{Rows in the same direction} \end{matrix}$$

The new start multiplies matrices for a purpose: Dependent columns are combinations of independent columns.

Please see the Table of Contents and the Preface: **math.mit.edu/everyone**

Chapter 2 : Elimination $A\boldsymbol{x} = \boldsymbol{b}$ reduces to an upper triangular system $U\boldsymbol{x} = \boldsymbol{c}$. Then back substitution for $\boldsymbol{x}$ is easy. The elimination steps go into a lower triangular matrix L with $\boldsymbol{A = LU}$. ILA5 gives a proof of this formula and LAFE adds a second explanation (using columns of L times rows of U, the fourth way to multiply matrices).

Chapter 3 : Vector spaces Both books introduce vector spaces, especially the **four fundamental subspaces** associated with A (m by n). The row space and column space have dimension r (the rank). The nullspaces of A and A^{T} have dimensions $n - r$ and $m - r$.



BIG PICTURE OF LINEAR ALGEBRA

row space $\perp$ nullspace column space of $A \perp$ nullspace of A^{T}

row rank = column rank = r

Elimination produces the matrix R for $A = CR$ in Chapter 1. LAFE explains the structure $R = \begin{bmatrix} I & F \end{bmatrix} P$ of this row echelon form—not seen elsewhere.

$\boldsymbol{A} = \boldsymbol{CR} = C\begin{bmatrix} I & F \end{bmatrix}P = \begin{bmatrix} C & CF \end{bmatrix}P =$ [**Indep** cols **Dependent** cols] **Permute** cols.

A related "magic factorization" is $\boldsymbol{A = CW^{-1}R^*}$, where the mixing matrix W is the r by r intersection of independent columns in C with independent rows of A in R^*.

Chapter 4 : Orthogonality The row space is orthogonal to the nullspace. This leads to the normal equation $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ for the least squares solution $\widehat{\boldsymbol{x}}$ to $A\boldsymbol{x} = \boldsymbol{b}$. The best example is fitting data points by the closest straight line.

If the columns of Q are orthonormal then $\boldsymbol{Q^{\mathrm{T}}Q = I}$. *These are very valuable matrices!* Constructing Q from the columns of A by "Gram-Schmidt" has become an essential algorithm. Orthogonalization is $A = QR$ with orthogonal Q and triangular R.

Chapter 5 : Determinants ILA5 approaches determinants by their properties, not their formulas. LAFE explains 3 by 3 determinants in detail. Either way leads to this hard-to-compute number with $\det AB = (\det A)(\det B)$. LAFE identifies $\det A$ as the volume of an n-dimensional tilted box. The simple proof of that volume formula was new to me.

Key point Each linear transformation $T(\boldsymbol{v})$ connects to a matrix multiplication $A\boldsymbol{v}$.

Chapter 6 : Eigenvalues An eigenvector $\boldsymbol{x}$ keeps the same direction when multiplied by $\boldsymbol{A}$. Then $A\boldsymbol{x} = \lambda\boldsymbol{x}$ and $(A - \lambda I)\boldsymbol{x} = \mathbf{0}$. Therefore $A - \lambda I$ has determinant zero. If the eigenvectors go into the columns of X, then $AX = X\Lambda$. The eigenvalues λ are on the diagonal of the matrix Λ.

Both ILA5 and LAFE show how eigenvalues lead to powers $\boldsymbol{A^n = X\Lambda^n X^{-1}}$. Both books solve differential equations $d\boldsymbol{u}/dt = A\boldsymbol{u}$. And both books emphasize symmetric matrices S (real eigenvalues λ with orthonormal eigenvectors in Q). Then $S = Q\Lambda Q^{\mathrm{T}} = S^{\mathrm{T}}$.

The best matrices of linear algebra are **symmetric positive definite matrices** (with positive eigenvalues). This topic beautifully connects eigenvalues to energy $\boldsymbol{x}^{\mathrm{T}} S\boldsymbol{x} > 0$.

Chapter 7 : Singular Values The SVD is highly important to linear algebra. It expresses every matrix as $A = U\Sigma V^{\mathrm{T}}$ with $U^{\mathrm{T}}U = I$ and $V^{\mathrm{T}}V = I$ and a diagonal matrix Σ of decreasing singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$. Then A multiplies an orthogonal basis in the row space to produce an orthogonal basis in the column space:

$$\boldsymbol{Av_i = \sigma_i u_i} \textbf{ instead of eigenvectors } \boldsymbol{Sq_i = \lambda_i q_i}. \text{ Two bases instead of one!}$$

The extra bonus from $A = U\Sigma V^{\mathrm{T}} = \boldsymbol{u}_1\sigma_1\boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{u}_2\sigma_2\boldsymbol{v}_2^{\mathrm{T}} + \cdots$ is that the first k terms give the rank k matrix A_k that comes closest to A. *Perfect for data science and image compression.* ILA5 gives examples in many fields. LAFE links to a remarkable website that compresses photographs supplied by the user. An excellent project for the class.

Chapter 7 of LAFE ends with an original essay on the victory of orthogonality:

orthogonal vectors, bases, subspaces, and matrices.

Chapter 8 : Learning from Data This very optional chapter of the new book explains deep learning—the creation of a function F that fits the known training data. The matrix weights are chosen to fit that data—then they give good results on unseen test data. The website **playground.tensorflow.org** shows the construction of F.

Chapters 8 to 12 of ILA5 (a longer book) show applications to graphs and networks and linear programming and Markov matrices and statistics.

Both books aim to explain the important ideas of linear algebra, *clearly and usefully.* A matrix becomes just as familiar as a derivative. To learn mathematics in the 21st century, this is the right goal.