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 $cv + dw$ fill a plane. Their dot products give length and angle: $v \cdot w = ||v|| \cdot ||w|| \cdot \cos \theta$. Then matrices multiply vectors in two ways: $Ax$ contains dot products of $x$ with the rows of $A$, and also $Ax$ is a **combination of the columns of $A$**. The first way is for hand computation. The second way is for understanding.

LAFE extends those linear combinations $Ax$ to matrix multiplication $AB$ in Chapter 1:

Column of $AB = A$ times column of $B =$ combination of the columns of $A$.

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 $A$)
3. All combinations of the columns (the column space of $A$)
4. $A = CR$ All columns of $A$ from the independent columns in $C$

$$A = \begin{bmatrix} 1 & 3 & 4 \\ 2 & 2 & 4 \\ 5 & 1 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 2 & 2 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} = CR$$

Column 3 = Columns 1 + 2

Column space = plane in $\mathbb{R}^3$

$A, C, R$ all have rank 2

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:

$$A = \begin{bmatrix} 2 & 4 & 10 \\ 3 & 6 & 15 \end{bmatrix} = \begin{bmatrix} 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 5 \end{bmatrix} = CR$$

Columns in the same direction

Rows in the same direction

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](http://math.mit.edu/everyone)
Chapter 2: Elimination  
\[ Ax = b \] reduces to an upper triangular system \( Ux = c \). Then back substitution for \( x \) is easy. The elimination steps go into a lower triangular matrix \( L \) with \( A = LU \). 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 \( m - r \).

\[
\begin{align*}
\text{row space} & \quad \text{column space} = \text{all } Ax \\
\text{null space} \quad \text{null space of } A^T \\
\text{row space} \perp \text{null space} & \quad \text{column space of } A \perp \text{null space of } A^T \\
\text{row rank} = \text{column rank} = r
\end{align*}
\]

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.

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

A related “magic factorization” is \( A = CW^{-1}R^* \), 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 \( A^T A \tilde{x} = A^T b \) for the least squares solution \( \tilde{x} \) to \( Ax = b \). The best example is fitting data points by the closest straight line.

If the columns of \( Q \) are orthonormal then \( Q^T Q = I \). 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 \( T(v) \) connects to a matrix multiplication \( Av \).

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

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

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

Chapter 7: Singular Values
The SVD is highly important to linear algebra. It expresses every matrix as \( A = U\Sigma V^T \) with \( U^TU = I \) and \( V^TV = I \) and a diagonal matrix \( \Sigma \) 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:

\[
Av_i = \sigma_i u_i \text{ instead of eigenvectors } S q_i = \lambda_i q_i. \text{ Two bases instead of one!}
\]

The extra bonus from \( A = U\Sigma V^T = u_1\sigma_1 v_1^T + u_2\sigma_2 v_2^T + \cdots \) 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.