The Four Fundamental Subspaces: 4 Lines

Gilbert Strang, Massachusetts Institute of Technology

1. Introduction. The expression "Four Fundamental Subspaces" has become familiar to thousands of linear algebra students. Those subspaces are the column space and the nullspace of A and A^{T}. They lift the understanding of $Ax = b$ to a higher level—a subspace level. The first step sees Ax (matrix times vector) as a *combination of the columns* of A. Those vectors Ax fill the column space $C(A)$. When we move from one combination to *all* combinations (by allowing every x), a subspace appears. $Ax = b$ has a solution exactly when b is in the column space of A.

The next section of this note will introduce all four subspaces. They are connected by the *Fundamental Theorem of Linear Algebra*. A perceptive reader may recognize the Singular Value Decomposition, when Part 3 of this theorem provides perfect bases for the four subspaces. The three parts are well separated in a linear algebra course! The first part goes as far as the *dimensions* of the subspaces, using the rank. The second part is their *orthogonality*—two subspaces in $\mathbf{R}^n$ and two in $\mathbf{R}^m$. The third part needs eigenvalues and eigenvectors of $A^{\mathrm{T}}A$ to find the best bases. Figure 1 will show the "big picture" of linear algebra, with the four bases added in Figure 2.

The main purpose of this paper is to see that theorem in action. We choose a matrix of rank one, $A = xy^{\mathrm{T}}$. When $m = n = 2$, all four fundamental subspaces are lines in $\mathbf{R}^2$. The big picture is particularly clear, and some would say the four lines are trivial. But the angle between x and y decides the eigenvalues of A and its Jordan form—those go beyond the Fundamental Theorem. We are seeing the orthogonal geometry that comes from singular vectors and the skew geometry that comes from eigenvectors. One leads to singular values and the other leads to eigenvalues.

Examples are amazingly powerful. I hope this family of 2 by 2 matrices fills a space between working with a specific numerical example and an arbitrary matrix.

2. The Four Subspaces. Figure 1 shows the fundamental subspaces for an m by n matrix of rank r. It is useful to fix ideas with a 3 by 4 matrix of rank 2:

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

That matrix is in row reduced echelon form and it shows what elimination can accomplish. The column space of A and the nullspace of A^{T} have very simple bases:

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \text{ span } C(A), \quad \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \text{ spans } N(A^{\mathrm{T}}).$$

After transposing, the first two rows of A are a basis for the row space—and they also tell us a basis for the nullspace:

$$\begin{bmatrix} 1 \\ 0 \\ 2 \\ 3 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ 1 \\ 4 \\ 5 \end{bmatrix} \text{ span } C(A^{\mathrm{T}}), \begin{bmatrix} -2 \\ -4 \\ 1 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} -3 \\ -5 \\ 0 \\ 1 \end{bmatrix} \text{ span } N(A).$$

The last two vectors are orthogonal to the first two. But these are not orthogonal bases. Elimination is enough to give Part 1 of the Fundamental Theorem:

Part 1
The column space and row space have equal dimension $r = $ rank
The nullspace $N(A)$ has dimension $n - r$, $N(A^{\mathrm{T}})$ has dimension $m - r$

That counting of basis vectors is obvious for the row reduced $\mathsf{rref}(A)$. This matrix has r nonzero rows and r pivot columns. The proof of Part 1 is in the reversibility of every elimination step—to confirm that linear independence and dimension are not changed.
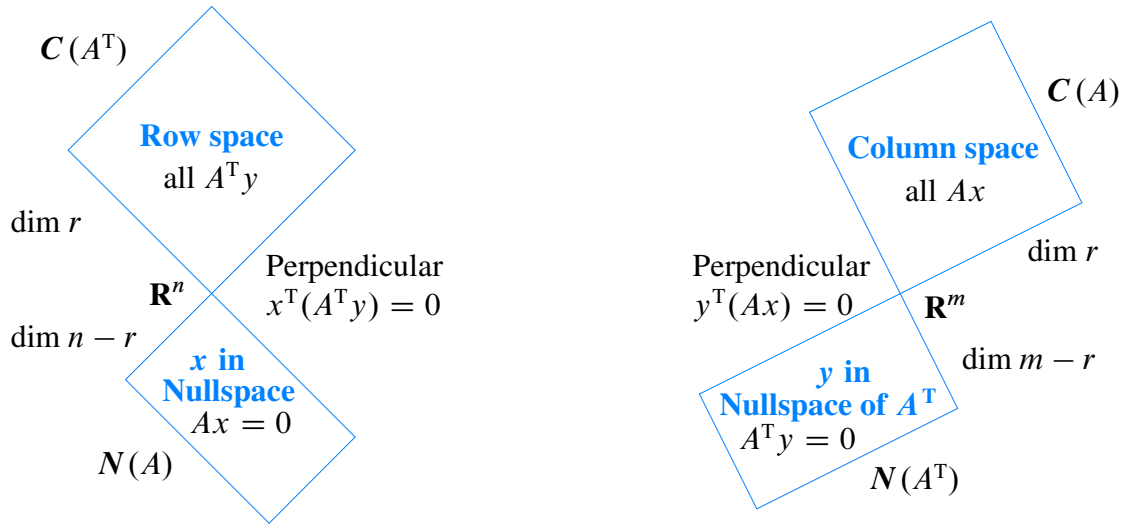

Figure 1: Dimensions and orthogonality for any m by n matrix A of rank r.

Part 2 says that the row space and nullspace are *orthogonal complements*. The orthogonality comes directly from the equation $Ax = 0$. Each x in the nullspace is orthogonal to each row:

$$Ax = 0 \qquad \begin{bmatrix} (\text{row } 1) \\ \cdots \\ (\text{row } m) \end{bmatrix} \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} 0 \\ \cdot\cdot \\ 0 \end{bmatrix} \begin{matrix} \longleftarrow x \text{ is orthogonal to row } 1 \\ \\ \longleftarrow x \text{ is orthogonal to row } m \end{matrix}$$

The dimensions of $C(A^{\mathrm{T}})$ and $N(A)$ add to n. Every vector in $\mathbf{R}^n$ is accounted for, by separating x into $x_{\mathsf{row}} + x_{\mathsf{null}}$.

For the 90° angle on the right side of Figure 1, change A to A^{T}. Every vector $b = Ax$ in the column space is orthogonal to every solution of $A^{\mathrm{T}}y = 0$. More briefly, $(Ax)^{\mathrm{T}}y = x^{\mathrm{T}}(A^{\mathrm{T}}y) = 0$.

Part 2
$C(A^{\mathrm{T}}) = N(A)^{\perp}$ Orthogonal complements in $\mathbf{R}^n$
$N(A^{\mathrm{T}}) = C(A)^{\perp}$ Orthogonal complements in $\mathbf{R}^m$

Part 3 of the Fundamental Theorem creates orthonormal bases for the four subspaces. More than that, the matrix is diagonal with respect to those bases $u_1, \ldots, u_n$ and $v_1, \ldots, v_m$. From row space to column space this is $Av_i = \sigma_i u_i$ for $i = 1, \ldots, r$. The other basis vectors are in the nullspaces: $Av_i = 0$ and $A^{\mathrm{T}}u_i = 0$ for $i > r$. When the u's and v's are columns of orthogonal matrices U and V, we have the Singular Value Decomposition $A = U\Sigma V^{\mathrm{T}}$:

$$\textbf{Part 3} \qquad \boldsymbol{AV} = A \begin{bmatrix} v_1 \cdot\cdot v_r \cdot\cdot v_n \end{bmatrix} = \begin{bmatrix} u_1 \cdot\cdot u_r \cdot\cdot u_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} = \boldsymbol{U\Sigma}.$$

The $v's$ are orthonormal eigenvectors of $A^{\mathrm{T}}A$, with eigenvalue $\sigma_i^2 \geq 0$. Then the eigenvector matrix V diagonalizes $A^{\mathrm{T}}A = (V\Sigma^{\mathrm{T}}U^{\mathrm{T}})(U\Sigma V^{\mathrm{T}}) = V(\Sigma^{\mathrm{T}}\Sigma)V^{\mathrm{T}}$. Similarly U diagonalizes AA^{T}.

When matrices are not symmetric or square, it is $A^{\mathrm{T}}A$ and AA^{T} that make things right.

This summary is completed by one more matrix: the **pseudoinverse**. This matrix A^+ inverts A where that is possible, from column space back to row space. It has the same nullspace as A^{T}. It gives the shortest solution to $Ax = b$, because A^+b is the particular solution in the row space: $AA^+b = b$. Every matrix is invertible from row space to column space, and A^+ provides the inverse:

$$\textbf{Pseudoinverse} \qquad A^+u_i = \frac{v_i}{\sigma_i} \quad \text{for} \quad i = 1, \ldots, r.$$

row space $\quad$ $Av_1 = \sigma_1 u_1$ $\quad$ column space
$A^+u_1 = v_1/\sigma_1$
v_1 $\quad$ u_1
v_2 $\quad$ u_2
$Av_2 = \sigma_2 u_2$
$A^+u_2 = v_2/\sigma_2$
v_4
v_3 $\quad$ u_3
$Av_3 = 0$
$Av_4 = 0$
$A^+u_3 = 0 = A^{\mathrm{T}}u_3$
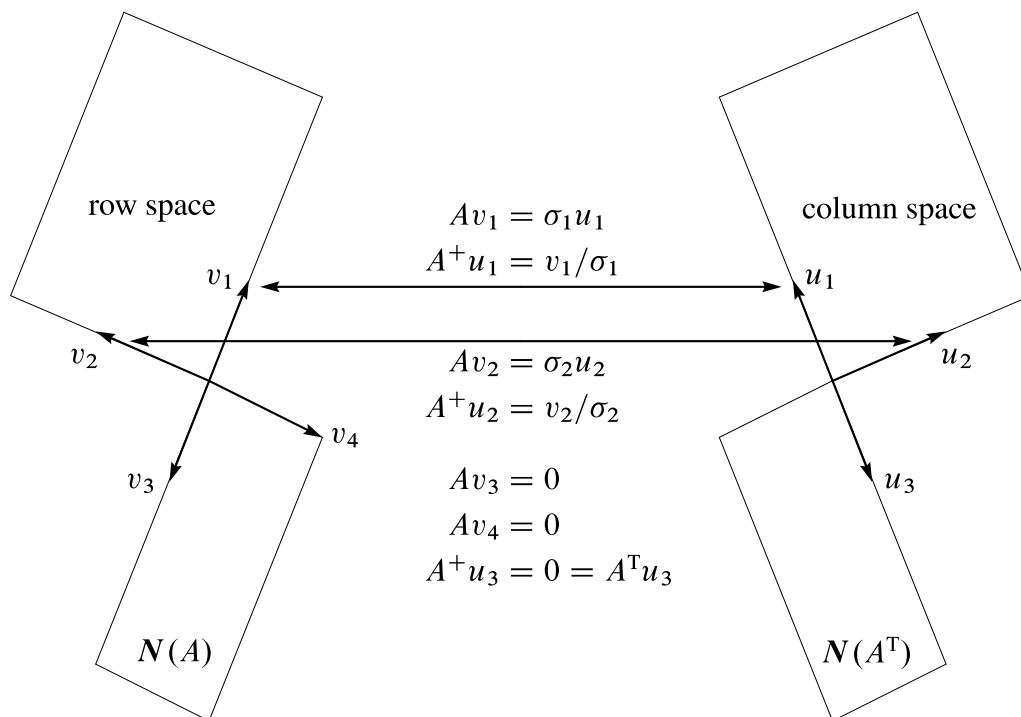$N(A)$ $\quad$ $N(A^{\mathrm{T}})$

Figure 2: Orthonormal bases that diagonalize A (3 by 4) and A^+ (4 by 3).

Figure 2 shows the four subspaces with orthonormal bases and the action of A and A^+. The product A^+A is the orthogonal projection of $\mathbf{R}^n$ onto the row space—as near to the identity matrix as possible. Certainly A^+ is A^{-1} when that inverse exists.

3. Matrices of Rank One. Our goal is a full understanding of rank one matrices $A = xy^{\mathrm{T}}$. The columns of A are multiples of x, so the column space $C(A)$ is a line. The rows of A are multiples of y^{T}, so the row space $C(A^{\mathrm{T}})$ is the line through y (column vector convention). Let x and y be unit vectors to make the scaling attractive. Since all the action concerns those two vectors, we stay in $\mathbf{R}^2$:

$$A = xy^{\mathrm{T}} = \begin{bmatrix} x_1 y_1 & x_1 y_2 \\ x_2 y_1 & x_2 y_2 \end{bmatrix}. \quad \text{The trace is } x_1 y_1 + x_2 y_2 = x^{\mathrm{T}} y.$$

The nullspace of A is the line orthogonal to y. It is in the direction of $y^\perp = (y_2, -y_1)$. The algebra gives $Ay^\perp = (xy^{\mathrm{T}})y^\perp = 0$ and the geometry is on the left side of Figure 3. The good basis vectors are y and $y^\perp$. On the right side, the bases for the column space of A and the nullspace of $A^{\mathrm{T}} = yx^{\mathrm{T}}$ are the orthogonal unit vectors x and $x^\perp$.

Figure 3: The fundamental subspaces for $A = xy^{\mathrm{T}}$ are four lines in $\mathbf{R}^2$.

4. Eigenvalues of xy^{T}. The eigenvalues of A were not mentioned in the Fundamental Theorem. Eigenvectors are not normally orthogonal. They belong to the column space and the nullspace, *not* a natural pair of subspaces. One subspace is in $\mathbf{R}^m$, one is in $\mathbf{R}^n$, and they are comparable (but usually not orthogonal) only when $m = n$. The eigenvectors of the singular 2 by 2 matrix $A = xy^{\mathrm{T}}$ are x and $y^\perp$:

Eigenvectors $\quad Ax = (xy^{\mathrm{T}})x = x(y^{\mathrm{T}}x) \quad$ and $\quad Ay^\perp = (xy^{\mathrm{T}})y^\perp = 0.$

The new and crucial number is that first eigenvalue $\lambda_1 = y^{\mathrm{T}}x = \cos\theta$. This is the trace since $\lambda_2 = 0$. The angle θ between row space and column space decides the orientation in Figure 3. The extreme cases $\theta = 0$ and $\theta = \pi/2$ produce matrices of the best kind and the worst kind:

Best $\quad \cos\theta = 1 \quad$ when $x = y$. $\quad$ Then $A = xx^{\mathrm{T}} \quad$ is symmetric with $\quad \lambda = 1, 0$

Worst $\quad \cos\theta = 0 \quad$ when $x = y^\perp$. $\quad$ Then $A = y^\perp y^{\mathrm{T}} \quad$ has trace zero with $\quad \lambda = 0, 0$

"Worst" is a short form of "nondiagonalizable". The eigenvalue $\lambda = 0$ is repeated and the two eigenvectors x and $y^\perp$ coincide when the trace $y^{\mathrm{T}}x$ is zero. At that point $A = xy^{\mathrm{T}}$

cannot be similar to the diagonal matrix of its eigenvalues (because this will be the zero matrix). The right choice of $Q^{-1}AQ$ will produce the Jordan form in this extreme case when x and y are orthonormal:

$$J = \begin{bmatrix} x^{\mathrm{T}} \\ y^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} xy^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} x & y \end{bmatrix} = \begin{bmatrix} x^{\mathrm{T}} \\ y^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} 0 & x \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

Jordan chose the best basis (x and y) to put xy^{T} in that famous form, with an off-diagonal 1 to signal a missing eigenvector. The SVD will choose *two different orthonormal bases* to put xy^{T} in its diagonal form Σ.

5. Factorizations of $A = xy^{\mathrm{T}}$. By bringing together three important ways to factor this matrix, you can see the end result of each approach and how that goal is reached. We still have $\|x\| = \|y\| = 1$ and $y^{\mathrm{T}}x = \cos\theta$. The end results are Σ, Λ, and T.

A. Singular Value Decomposition $\qquad U^{\mathrm{T}}AV = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \Sigma$

B. Diagonalization by eigenvectors $\qquad S^{-1}AS = \begin{bmatrix} \cos\theta & 0 \\ 0 & 0 \end{bmatrix} = \Lambda$

C. Orthogonal triangularization $\qquad Q^{\mathrm{T}}AQ = \begin{bmatrix} \cos\theta & \sin\theta \\ 0 & 0 \end{bmatrix} = T$

The columns of U, V, S, and Q will be x, y, $x^{\perp}$, and $y^{\perp}$. They come in different orders!
A. In the SVD, the columns of U and V are orthonormal bases for the four subspaces. Figure 3 shows $u_1 = x$ in the column space and $v_1 = y$ in the row space. Then $Ay = (xy^{\mathrm{T}})y$ correctly gives x with $\sigma_1 = 1$. The nullspace bases are $u_2 = x^{\perp}$ and $v_2 = y^{\perp}$. Notice the different bases in U and V, from the reversal of x and y:

$$U^{\mathrm{T}}AV = \begin{bmatrix} x & x^{\perp} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} xy^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} y & y^{\perp} \end{bmatrix} = \begin{bmatrix} x & x^{\perp} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} x & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \Sigma$$

The pseudoinverse of xy^{T} is yx^{T}. The norm of A is $\sigma_1 = 1$.
B. In diagonalization, the eigenvectors of $A = xy^{\mathrm{T}}$ are x and $y^{\perp}$. Those are the columns of the eigenvector matrix S, and its determinant is $y^{\mathrm{T}}x = \cos\theta$. The eigenvectors of $A^{\mathrm{T}} = yx^{\mathrm{T}}$ are y and $x^{\perp}$, which go into the rows of S^{-1} (after division by $\cos\theta$).

$$S^{-1}AS = \frac{1}{\cos\theta} \begin{bmatrix} y & x^{\perp} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} xy^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} x & y^{\perp} \end{bmatrix} = \begin{bmatrix} y & 0 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} x & 0 \end{bmatrix} = \begin{bmatrix} \cos\theta & 0 \\ 0 & 0 \end{bmatrix} = \Lambda$$

This diagonalization fails when $\cos\theta = 0$ and S is singular. The Jordan form jumps from Λ to J, as that off-diagonal 1 suddenly appears.
C. One of the many useful discoveries of Isaac Schur is that every square matrix is unitarily similar to a triangular matrix:

$$Q^{*}AQ = T \quad \text{with} \quad Q^{*}Q = I.$$

His construction starts with the unit eigenvector x in the first column of Q. In our 2 by 2 case, the construction ends immediately with $x^\perp$ in the second column:

$$Q^\mathrm{T} A Q = \begin{bmatrix} x^\mathrm{T} \\ x^{\perp\mathrm{T}} \end{bmatrix} \begin{bmatrix} xy^\mathrm{T} \end{bmatrix} \begin{bmatrix} x & x^\perp \end{bmatrix} = \begin{bmatrix} y^\mathrm{T} \\ 0^\mathrm{T} \end{bmatrix} \begin{bmatrix} x & x^\perp \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ 0 & 0 \end{bmatrix} = T$$

This triangular matrix T still has norm 1, since Q is unitary. Numerically T is far more stable than the diagonal form Λ. In fact T survives in the limit $\cos\theta = 0$ of coincident eigenvectors, when it becomes J.

Note The triangular form T is not so widely used, but it gives an elementary proof of a seemingly obvious fact: A random small perturbation of any square matrix is almost sure to produce distinct eigenvalues. What is the best proof?

More controversially, I wonder if Schur can be regarded as the greatest linear algebraist of all time?

Summary The four fundamental subspaces, coming from $A = xy^\mathrm{T}$ and from $A^\mathrm{T} = yx^\mathrm{T}$, are four lines in $\mathbf{R}^2$. Their directions are given by $x, x^\perp, y$, and $y^\perp$. The eigenvectors of A and A^T are the same four vectors. But there is a crucial crossover in the pictures of Figures 1-2-3. The eigenvectors of A lie in its column space and nullspace, not a natural pair. The dimensions of the spaces add to $n = 2$, but the spaces are not orthogonal and they could even coincide.

The better picture is the orthogonal one that leads to the SVD.

References

These are among the textbooks that present the four subspaces.

1. David Lay, *Linear Algebra and Its Applications*, Third edition, Addison-Wesley (2003).

2. Peter Olver and Chehrzad Shakiban, *Applied Linear Algebra*, Pearson Prentice-Hall (2006).

3. Theodore Shifrin and Malcolm Adams, *Linear Algebra: A Geometric Approach*, Freeman (2001).

4. Gilbert Strang, *Linear Algebra and Its Applications*, Fourth edition, Cengage (previously Brooks/Cole) (2006).

5. Gilbert Strang, *Introduction to Linear Algebra*, Third edition, Wellesley-Cambridge Press (2003).