Chapter 8

Applications

8.1 Matrices in Engineering

This section will show how engineering problems produce symmetric matrices K (often K is positive definite). The "linear algebra reason" for symmetry and positive definiteness is their form $K = A^{\mathrm{T}}A$ and $K = A^{\mathrm{T}}CA$. The "physical reason" is that the expression $\frac{1}{2}\boldsymbol{u}^{\mathrm{T}}K\boldsymbol{u}$ represents *energy*—and energy is never negative. The matrix C, often diagonal, contains positive physical constants like conductance or stiffness or diffusivity.

Our first examples come from mechanical and civil and aeronautical engineering. K is the ***stiffness matrix***, and $K^{-1}f$ is the structure's response to forces f from outside. Section 8.2 turns to electrical engineering—the matrices come from networks and circuits. The exercises involve chemical engineering and I could go on! Economics and management and engineering design come later in this chapter (there the key is optimization).

Engineering leads to linear algebra in two ways, directly and indirectly:

Direct way The physical problem has only a finite number of pieces. The laws connecting their position or velocity are *linear* (movement is not too big or too fast). The laws are expressed by *matrix equations.*

Indirect way The physical system is "continuous". Instead of individual masses, the mass density and the forces and the velocities are functions of x or x, y or x, y, z. The laws are expressed by *differential equations.* ***To find accurate solutions we approximate by finite difference equations or finite element equations.***

Both ways produce matrix equations and linear algebra. I really believe that you cannot do modern engineering without matrices.

Here we present equilibrium equations $K\boldsymbol{u} = f$. With motion, $Md^2\boldsymbol{u}/dt^2 + K\boldsymbol{u} = f$ becomes dynamic. Then we use eigenvalues from $K\boldsymbol{x} = \lambda M\boldsymbol{x}$, or finite differences.

Before explaining the physical examples, may I write down the matrices? The tridiagonal K_0 appears many times in this textbook. Now we will see its applications. These matrices are all symmetric, and the first four are positive definite:

$$K_0 = A_0^{\mathrm{T}}A_0 = \begin{bmatrix} 2 & -1 & \\ -1 & 2 & -1 \\ & -1 & \mathbf{2} \end{bmatrix} \qquad A_0^{\mathrm{T}}C_0A_0 = \begin{bmatrix} c_1+c_2 & -c_2 & \\ -c_2 & c_2+c_3 & -c_3 \\ & -c_3 & c_3+c_4 \end{bmatrix}$$

Fixed-fixed **Spring constants included**

$$K_1 = A_1^{\mathrm{T}}A_1 = \begin{bmatrix} 2 & -1 & \\ -1 & 2 & -1 \\ & -1 & \mathbf{1} \end{bmatrix} \qquad A_1^{\mathrm{T}}C_1A_1 = \begin{bmatrix} c_1+c_2 & -c_2 & \\ -c_2 & c_2+c_3 & -c_3 \\ & -c_3 & c_3 \end{bmatrix}$$

Fixed-free **Spring constants included**

$$K_{\textbf{singular}} = \begin{bmatrix} \mathbf{1} & -1 & \\ -1 & 2 & -1 \\ & -1 & \mathbf{1} \end{bmatrix} \qquad K_{\textbf{circular}} = \begin{bmatrix} 2 & -1 & \mathbf{-1} \\ -1 & 2 & -1 \\ \mathbf{-1} & -1 & 2 \end{bmatrix}$$

Free-free

The matrices $K_0, K_1, K_{\text{singular}}$, and K_{circular} have $C = I$ for simplicity. This means that all the "spring constants" are $c_i = 1$. We included $A_0^{\mathrm{T}}C_0A_0$ and $A_1^{\mathrm{T}}C_1A_1$ to show how the spring constants enter the matrix (without changing its positive definiteness). Our first goal is to show where these stiffness matrices come from.

A Line of Springs

Figure 8.1 shows three masses m_1, m_2, m_3 connected by a line of springs. One case has four springs, with top and bottom fixed. The fixed-free case has only three springs; the lowest mass hangs freely. The **fixed-fixed** problem will lead to K_0 and $A_0^{\mathrm{T}}C_0A_0$. The **fixed-free** problem will lead to K_1 and $A_1^{\mathrm{T}}C_1A_1$. A **free-free** problem, with no support at either end, produces the matrix K_{singular}.

We want equations for the mass movements $\boldsymbol{u}$ and the tensions (or compressions) $\boldsymbol{y}$:

$$\begin{aligned} \boldsymbol{u} &= (u_1, u_2, u_3) = \textbf{movements of the masses} \text{ (down or up)} \\ \boldsymbol{y} &= (y_1, y_2, y_3, y_4) \text{ or } (y_1, y_2, y_3) = \textbf{tensions in the springs} \end{aligned}$$

When a mass moves downward, its displacement is positive ($u_i > 0$). For the springs, tension is positive and compression is negative ($y_i < 0$). In tension, the spring is stretched so it pulls the masses inward. Each spring is controlled by its own Hooke's Law $y = c\,e$: (*stretching force*) = (*spring constant*) times (*stretching distance*).

Our job is to link these one-spring equations $y = ce$ into a vector equation $Ku = f$ for the whole system. The force vector f comes from gravity. The gravitational constant g will multiply each mass to produce forces $f = (m_1g, m_2g, m_3g)$.

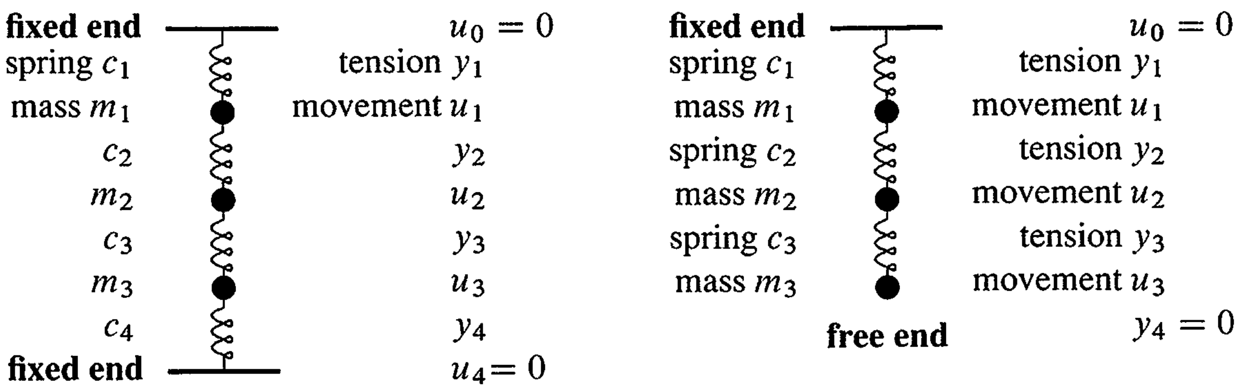


Figure 8.1: Lines of springs and masses: **fixed-fixed** and **fixed-free** ends.

The real problem is to find the stiffness matrix (**fixed-fixed** and **fixed-free**). The best way to create K is in three steps, not one. Instead of connecting the movements $\boldsymbol{u}_i$ directly to the forces, it is much better to connect each vector to the next in this list:

$$\begin{array}{lclcl} u &=& \textit{Movements} \text{ of } n \text{ masses} &=& (u_1,\ldots,u_n) \\ e &=& \textit{Elongations} \text{ of } m \text{ springs} &=& (e_1,\ldots,e_m) \\ y &=& \textit{Internal forces} \text{ in } m \text{ springs} &=& (y_1,\ldots,y_m) \\ f &=& \textit{External forces} \text{ on } n \text{ masses} &=& (f_1,\ldots,f_n) \end{array}$$

The framework that connects $\boldsymbol{u}$ to $\boldsymbol{e}$ to $\boldsymbol{y}$ to $\boldsymbol{f}$ looks like this:

$$\begin{tikzcd} \boxed{u} \arrow[d, "A"'] & \boxed{f} \\ \boxed{e} \arrow[r, "C"] & \boxed{y} \arrow[u, "A^{\mathrm{T}}"'] \end{tikzcd}$$

$$\begin{array}{ll} e = Au & A \text{ is } m \text{ by } n \\ y = Ce & C \text{ is } m \text{ by } m \\ f = A^{\mathrm{T}}y & A^{\mathrm{T}} \text{ is } n \text{ by } m \end{array}$$

We will write down the matrices A and C and A^{T} for the two examples, first with fixed ends and then with the lower end free. Forgive the simplicity of these matrices, it is their form that is so important. Especially the appearance of A together with A^{T}.

The *elongation* $\boldsymbol{e}$ is the stretching distance—how far the springs are extended. Originally there is no stretching—the system is lying on a table. When it becomes vertical and upright, gravity acts. The masses move down by distances u_1, u_2, u_3. Each spring is stretched or compressed by $e_i = u_i - u_{i-1}$, *the difference in displacements of its ends*:

	First spring:	$e_1 = u_1$	(the top is fixed so $u_0 = 0$)
Stretching of	Second spring:	$e_2 = u_2 - u_1$	
each spring	Third spring:	$e_3 = u_3 - u_2$	
	Fourth spring:	$e_4 = \quad - u_3$	(the bottom is fixed so $u_4 = 0$)

If both ends move the same distance, that spring is not stretched: $u_i = u_{i-1}$ and $e_i = 0$. The matrix in those four equations is a 4 by 3 *difference matrix* A, and $\boldsymbol{e} = A\boldsymbol{u}$:

$$\begin{array}{c}\textbf{Stretching}\\ \textbf{distances}\\ \textbf{(elongations)}\end{array} \quad \boldsymbol{e} = A\boldsymbol{u} \quad \text{is} \quad \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}. \tag{1}$$

The next equation $\boldsymbol{y} = C\boldsymbol{e}$ connects spring elongation $\boldsymbol{e}$ with spring tension $\boldsymbol{y}$. *This is Hooke's Law $y_i = c_i e_i$ for each separate spring.* It is the "constitutive law" that depends on the material in the spring. A soft spring has small c, so a moderate force y can produce a large stretching e. Hooke's linear law is nearly exact for real springs, before they are overstretched and the material becomes plastic.

Since each spring has its own law, the matrix in $\boldsymbol{y} = C\boldsymbol{e}$ is a diagonal matrix C:

$$\begin{array}{c}\textbf{Hooke's}\\ \textbf{Law}\\ y = Ce\end{array} \quad \begin{array}{rcl} y_1 &=& c_1e_1 \\ y_2 &=& c_2e_2 \\ y_3 &=& c_3e_3 \\ y_4 &=& c_4e_4 \end{array} \quad \text{is} \quad \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} c_1 & & & \\ & c_2 & & \\ & & c_3 & \\ & & & c_4 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} \tag{2}$$

Combining $\boldsymbol{e} = A\boldsymbol{u}$ with $\boldsymbol{y} = C\boldsymbol{e}$, the spring forces are $\boldsymbol{y} = CA\boldsymbol{u}$.

Finally comes the *balance equation*, the most fundamental law of applied mathematics. The internal forces from the springs balance the external forces on the masses. Each mass is pulled or pushed by the spring force y_j above it. From below it feels the spring force y_{j+1} plus f_j from gravity. Thus $y_j = y_{j+1} + f_j$ or $f_j = y_j - y_{j+1}$:

$$\begin{array}{c}\textbf{Force}\\ \textbf{balance}\\ f = A^{\mathrm{T}}y\end{array} \quad \begin{array}{rcl} f_1 &=& y_1 - y_2 \\ f_2 &=& y_2 - y_3 \\ f_3 &=& y_3 - y_4 \end{array} \quad \text{is} \quad \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \tag{3}$$

That matrix is** A^{T}. The equationfor balance offorces is $\boldsymbol{f} = A^{\mathrm{T}}\boldsymbol{y}$.* Nature transposes the rows and columns of the $\boldsymbol{e} - \boldsymbol{u}$ matrix to produce the $\boldsymbol{f} - \boldsymbol{y}$ matrix. This is the beauty of the framework, that A^{T} appears along with A. The three equations combine into $K\boldsymbol{u} = \boldsymbol{f}$, where the ***stiffness matrix is $K = A^{\mathrm{T}}CA$:

$$\left\{ \begin{array}{rcl} e &=& Au \\ y &=& Ce \\ f &=& A^{\mathrm{T}}y \end{array} \right\} \quad \text{combine into} \quad A^{\mathrm{T}}CAu = f \quad \text{or} \quad Ku = f\,.$$

In the language of elasticity, $\boldsymbol{e} = A\boldsymbol{u}$ is the **kinematic** equation (for displacement). The force balance $\boldsymbol{f} = A^{\mathrm{T}}\boldsymbol{y}$ is the **static** equation (for equilibrium). The **constitutive law** is $\boldsymbol{y} = C\boldsymbol{e}$ (from the material). Then $A^{\mathrm{T}}CA$ is n by n = (n by m)(m by m)(m by n).

Finite element programs spend major effort on assembling $K = A^{\mathrm{T}}CA$ from thousands of smaller pieces. We find K for four springs (**fixed-fixed**) by multiplying A^{T} times CA:

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} c_1 & 0 & 0 \\ -c_2 & c_2 & 0 \\ 0 & -c_3 & c_3 \\ 0 & 0 & -c_4 \end{bmatrix} = \begin{bmatrix} c_1 + c_2 & -c_2 & 0 \\ -c_2 & c_2 + c_3 & -c_3 \\ 0 & -c_3 & c_3 + c_4 \end{bmatrix}$$

If all springs are identical, with $c_1 = c_2 = c_3 = c_4 = 1$, then $C = I$. The stiffness matrix reduces to $A^{\mathrm{T}}A$. It becomes the special $-1, 2, -1$ matrix:

$$\textbf{With } C = I \qquad K_0 = A_0^{\mathrm{T}}A_0 = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}. \tag{4}$$

Note the difference between $A^{\mathrm{T}}A$ from engineering and LL^{T} from linear algebra. The matrix A from four springs is 4 by 3. The triangular matrix L from elimination is square. The stiffness matrix K is assembled from $A^{\mathrm{T}}A$, and then broken up into LL^{T}. One step is applied mathematics, the other is computational mathematics. Each K is built from rectangular matrices and factored into square matrices.

May I list some properties of $K = A^{\mathrm{T}}CA$? You know almost all of them:

1. K is **tridiagonal**, because mass 3 is not connected to mass 1.
2. K is **symmetric**, because C is symmetric and A^{T} comes with A.
3. K is **positive definite**, because $c_i > 0$ and A has **independent columns**.
4. K^{-1} is a full matrix in equation (5) with **all positive entries**.

That last property leads to an important fact about $\boldsymbol{u} = K^{-1}\boldsymbol{f}$: *If all forces act downwards* $(f_j > 0)$ *then all movements are downwards* $(u_j > 0)$. Notice that "positiveness" is different from "positive definiteness". Here K^{-1} is positive (K is not). Both K and K^{-1} are positive definite.

Example 1 Suppose all $c_i = c$ and $m_j = m$. Find the movements $\boldsymbol{u}$ and tensions $\boldsymbol{y}$.

All springs are the same and all masses are the same. But all movements and elongations and tensions will *not* be the same. K^{-1} includes $\frac{1}{c}$ because $A^{\mathrm{T}}CA$ includes c:

$$\boldsymbol{u} = K^{-1}\boldsymbol{f} = \frac{1}{4c}\begin{bmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix}\begin{bmatrix} mg \\ mg \\ mg \end{bmatrix} = \frac{mg}{c}\begin{bmatrix} 3/2 \\ 2 \\ 3/2 \end{bmatrix} \tag{5}$$

The displacement u_2, for the mass in the middle, is greater than u_1 and u_3. The units are correct: the force mg divided by force per unit length c gives a length u. Then

$$\boldsymbol{e} = A\boldsymbol{u} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}\frac{mg}{c}\begin{bmatrix} \frac{3}{2} \\ 2 \\ \frac{3}{2} \end{bmatrix} = \frac{mg}{c}\begin{bmatrix} 3/2 \\ 1/2 \\ -1/2 \\ -3/2 \end{bmatrix}.$$

Those elongations add to zero because the ends of the line are fixed. (The sum $u_1 + (u_2 - u_1) + (u_3 - u_2) + (-u_3)$ is certainly zero.) For each spring force y_i we just multiply e_i by c. So y_1, y_2, y_3, y_4 are $\frac{3}{2}mg$, $\frac{1}{2}mg$, $-\frac{1}{2}mg$, $-\frac{3}{2}mg$. The upper two springs are stretched, the lower two springs are compressed.

Notice how $\boldsymbol{u}, \boldsymbol{e}, \boldsymbol{y}$ are computed in that order. We assembled $K = A^{\mathrm{T}}CA$ from rectangular matrices. To find $\boldsymbol{u} = K^{-1}\boldsymbol{f}$, we work with the whole matrix and not its three pieces! The rectangular matrices A and A^{T} do not have (two-sided) inverses.

Warning: *Normally you cannot write* $K^{-1} = A^{-1}C^{-1}(A^{\mathrm{T}})^{-1}$.

The three matrices are mixed together by $A^{\mathrm{T}}CA$, and they cannot easily be untangled. In general, $A^{\mathrm{T}}y = f$ has many solutions. And four equations $Au = e$ would usually have no solution with three unknowns. But $A^{\mathrm{T}}CA$ gives the correct solution to all three equations in the framework. Only when $m = n$ and the matrices are square can we go from $y = (A^{\mathrm{T}})^{-1}f$ to $e = C^{-1}y$ to $u = A^{-1}e$. We will see that now.

Fixed End and Free End

Remove the fourth spring. All matrices become 3 by 3. The pattern does not change! The matrix A loses its fourth row and (of course) A^{T} loses its fourth column. The new stiffness matrix K_1 becomes a product of square matrices:

$$A_1^{\mathrm{T}}C_1A_1 = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_1 & & \\ & c_2 & \\ & & c_3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}.$$

The missing column of A^{T} and row of A multiplied the missing c_4. So the quickest way to find the new $A^{\mathrm{T}}CA$ is to set $c_4 = 0$ in the old one:

FIXED FREE
$$K_1 = A_1^{\mathrm{T}}C_1A_1 = \begin{bmatrix} c_1 + c_2 & -c_2 & 0 \\ -c_2 & c_2 + c_3 & -c_3 \\ 0 & -c_3 & c_3 \end{bmatrix}. \qquad (6)$$

If $c_1 = c_2 = c_3 = 1$ and $C = I$, this is the $-1, 2, -1$ tridiagonal matrix, except the last entry is 1 instead of 2. The spring at the bottom is free.

Example 2 All $c_i = c$ and all $m_j = m$ in the fixed-free hanging line of springs. Then

$$K_1 = c\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \quad \text{and} \quad K_1^{-1} = \frac{1}{c}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}.$$

The forces mg from gravity are the same. But the movements change from the previous example because the stiffness matrix has changed:

$$u = K_1^{-1}f = \frac{1}{c}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}\begin{bmatrix} mg \\ mg \\ mg \end{bmatrix} = \frac{mg}{c}\begin{bmatrix} 3 \\ 5 \\ 6 \end{bmatrix}.$$

Those movements are greater in this fixed-free case. The number 3 appears in u_1 because all three masses are pulling the first spring down. The next mass moves by that 3 plus an additional 2 from the masses below it. The third mass drops even more $(3 + 2 + 1 = 6)$. The elongations $e = Au$ in the springs display those numbers $3, 2, 1$:

$$e = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}\frac{mg}{c}\begin{bmatrix} 3 \\ 5 \\ 6 \end{bmatrix} = \frac{mg}{c}\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}.$$

Multiplying by c, the forces y in the three springs are $3mg$ and $2mg$ and mg. And the special point of square matrices is that y can be found directly from f! The balance equation $A^{\mathrm{T}}y = f$ determines y immediately, because $m = n$ and A^{T} is square. We are allowed to write $(A^{\mathrm{T}}CA)^{-1} = A^{-1}C^{-1}(A^{\mathrm{T}})^{-1}$:

$$y = (A^{\mathrm{T}})^{-1}f \quad \text{is} \quad \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} mg \\ mg \\ mg \end{bmatrix} = \begin{bmatrix} 3mg \\ 2mg \\ 1mg \end{bmatrix}.$$

Two Free Ends: K is Singular

The first line of springs in Figure 8.2 is free at *both ends*. This means trouble (the whole line can move). The matrix A is 2 by 3, short and wide. Here is $e = Au$:

$$\textbf{FREE-FREE} \qquad \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} u_2 - u_1 \\ u_3 - u_2 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}. \qquad (7)$$

Now there is a nonzero solution to $Au = \mathbf{0}$. ***The masses can move with no stretching of the springs.*** The whole line can shift by $u = (1, 1, 1)$ and this leaves $e = (0, 0)$. A has *dependent columns* and the vector $(1, 1, 1)$ is in its nullspace:

$$Au = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \textbf{no stretching}. \qquad (8)$$

$Au = \mathbf{0}$ certainly leads to $A^{\mathrm{T}}CAu = \mathbf{0}$. So $A^{\mathrm{T}}CA$ is only *positive semidefinite*, without c_1 and c_4. The pivots will be c_2 and c_3 and *no third pivot*. The rank is only 2:

$$\begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} c_2 & \\ & c_3 \end{bmatrix} \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} c_2 & -c_2 & 0 \\ -c_2 & c_2 + c_3 & -c_3 \\ 0 & -c_3 & c_3 \end{bmatrix} \qquad (9)$$

Two eigenvalues will be positive but $x = (1, 1, 1)$ is an eigenvector for $\lambda = 0$. We can solve $A^{\mathrm{T}}CAu = f$ only for special vectors f. The forces have to add to $f_1 + f_2 + f_3 = 0$, or the whole line of springs (with both ends free) will take off like a rocket.

Circle of Springs

A third spring will complete the circle from mass 3 back to mass 1. This doesn't make K invertible—the new matrix is still singular. That stiffness matrix $K_{circular}$ is not tridiagonal, but it is symmetric (always) and *semidefinite*:

$$A^{\mathrm{T}}_{\textbf{circular}}A_{\textbf{circular}} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}. \qquad (10)$$

The only pivots are 2 and $\frac{3}{2}$. The eigenvalues are 3 and 3 and 0. The determinant is zero. The nullspace still contains $x = (1, 1, 1)$, when all the masses move together.

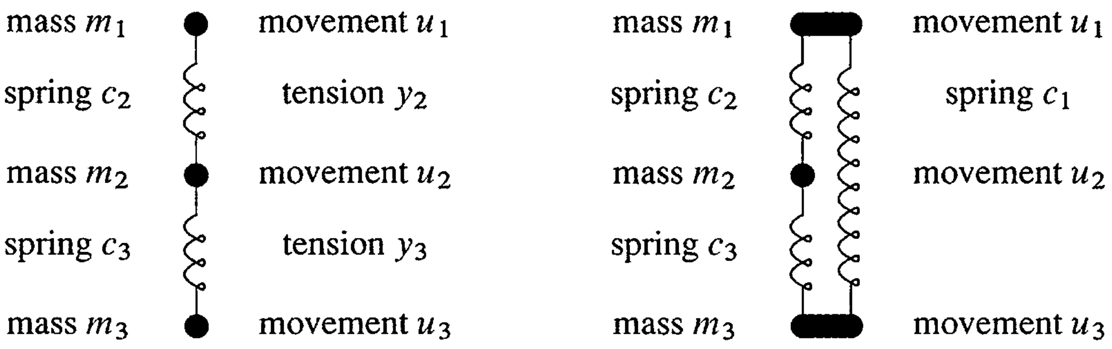


Figure 8.2: **Free-free ends**: A line of springs and a "circle" of springs: *Singular* K's. The masses can move without stretching the springs so $A\boldsymbol{u} = \mathbf{0}$ has nonzero solutions.

This movement vector $(1, 1, 1)$ is in the nullspace of $A_{\textbf{circular}}$ and $K_{\textbf{circular}}$, even after the diagonal matrix C of spring constants is included: the springs are not stretched.

$$(A^{\mathrm{T}}CA)_{\textbf{circular}} = \begin{bmatrix} c_1 + c_2 & -c_2 & -c_1 \\ -c_2 & c_2 + c_3 & -c_3 \\ -c_1 & -c_3 & c_3 + c_1 \end{bmatrix}. \tag{11}$$

Continuous Instead of Discrete

Matrix equations are discrete. Differential equations are continuous. We will see the differential equation that corresponds to the tridiagonal $-1, 2, -1$ matrix $A^{\mathrm{T}}A$. And it is a pleasure to see the boundary conditions that go with K_0 and K_1.

The matrices A and A^{T} correspond to the derivatives d/dx and $-d/dx$! Remember that $\boldsymbol{e} = A\boldsymbol{u}$ took differences $u_i - u_{i-1}$, and $f = A^{\mathrm{T}}\boldsymbol{y}$ took differences $y_i - y_{i+1}$. Now the springs are infinitesimally short, and those differences become derivatives:

$$\frac{u_i - u_{i-1}}{\Delta x} \text{ is like } \frac{du}{dx} \qquad \frac{y_i - y_{i+1}}{\Delta x} \text{ is like } -\frac{dy}{dx}$$

The factor Δx didn't appear earlier—we imagined the distance between masses was 1. To approximate a continuous solid bar, we take many more masses (smaller and closer). Let me jump to the three steps A, C, A^{T} in the continuous model, when there is stretching and Hooke's Law and force balance at every point x:

$$e(x) = Au = \frac{du}{dx} \qquad y(x) = c(x)e(x) \qquad A^{\mathrm{T}}y = -\frac{dy}{dx} = f(x)$$

Combining those equations into $A^{\mathrm{T}}CAu(x) = f(x)$, we have a differential equation not a matrix equation. The line of springs becomes an elastic bar:

$$\textbf{Solid Elastic Bar} \qquad A^{\mathrm{T}}CAu(x) = f(x) \quad \text{is} \quad -\frac{d}{dx}\left(c(x)\frac{du}{dx}\right) = f(x) \tag{12}$$

A^TA corresponds to a second derivative. A is a "difference matrix" and A^TA is a "second difference matrix". ***The matrix has*** $-1, 2, -1$ ***and the equation has*** $-d^2u/dx^2$:

$$-u_{i+1} + 2u_i - u_{i-1} \text{ is a \textbf{second difference}} \qquad -\frac{d^2u}{dx^2} \text{ is a \textbf{second derivative}.}$$

Now we see why this symmetric matrix is a favorite. When we meet a first derivative du/dx, we have three choices (*forward, backward, and centered differences*):

$$\frac{du}{dx} \simeq \frac{u(x+\Delta x)-u(x)}{\Delta x} \text{ or } \frac{u(x)-u(x-\Delta x)}{\Delta x} \text{ or } \frac{u(x+\Delta x)-u(x-\Delta x)}{2\Delta x}.$$

When we meet d^2u/dx^2, the natural choice is $u(x+\Delta x) - 2u(x) + u(x-\Delta x)$, divided by $(\Delta x)^2$. *Why reverse these signs to* $-1, 2, -1$? Because the positive definite matrix has $+2$ on the diagonal. First derivatives are *anti*symmetric; the transpose has a minus sign. So second differences are negative definite, and we change to $-d^2u/dx^2$.

We have moved from vectors to functions. Scientific computing moves the other way. It starts with a differential equation like (12). Sometimes there is a formula for the solution $u(x)$, more often not. In reality we *create* the discrete matrix K by approximating the continuous problem. Watch how the boundary conditions on u come in! By missing u_0 we treat it (correctly) as zero:

$$\begin{matrix}\textbf{FIXED}\\ \textbf{FIXED}\end{matrix} \quad Au = \frac{1}{\Delta x}\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \approx \frac{du}{dx} \quad \text{with} \quad \begin{matrix} \boldsymbol{u_0 = 0} \\ \boldsymbol{u_4 = 0} \end{matrix} \qquad (13)$$

Fixing the top end gives the boundary condition $u_0 = 0$. What about the free end, when the bar hangs in the air? Row 4 of A is gone and so is u_4. The boundary condition must come from A^T. It is the missing y_4 that we are treating (correctly) as zero:

$$\begin{matrix}\textbf{FIXED}\\ \textbf{FREE}\end{matrix} \quad A^Ty = \frac{1}{\Delta x}\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \approx -\frac{dy}{dx} \quad \text{with} \quad \begin{matrix} \boldsymbol{u_0 = 0} \\ \boldsymbol{y_4 = 0} \end{matrix} \qquad (14)$$

The boundary condition $y_4 = 0$ *at the free end becomes* $du/dx = 0$, since $y = Au$ corresponds to du/dx. The force balance $A^Ty = f$ at that end (in the air) is $0 = 0$. The last row of $K_1u = f$ has entries $-1, 1$ to reflect this condition $du/dx = 0$.

May I summarize this section? I hope this example will help you turn calculus into linear algebra, replacing differential equations by difference equations. If your step Δx is small enough, you will have a totally satisfactory solution.

$$\textbf{The equation is} \quad -\frac{d}{dx}\left(c(x)\frac{du}{dx}\right) = f(x) \textbf{ with } u(0) = 0 \textbf{ and } \left[u(1) \text{ or } \frac{du}{dx}(1)\right] = 0$$

Divide the bar into N pieces of length Δx. Replace du/dx by Au and $-dy/dx$ by A^Ty. Now A and A^T include $1/\Delta x$. The end conditions are $u_0 = 0$ and $[u_N = 0$ or $y_N = 0]$.

The three steps $-d/dx$ and $c(x)$ and d/dx correspond to A^{T} and C and A:

$$f = A^{\mathrm{T}}y \quad \text{and} \quad y = Ce \quad \text{and} \quad e = Au \quad \text{give} \quad A^{\mathrm{T}}CAu = f.$$

This is a fundamental example in computational science and engineering. Our book concentrates on Step 3 in that process (linear algebra). Now we have taken Step 2.

1. Model the problem by a differential equation
2. Discretize the differential equation to a difference equation
3. Understand and solve the difference equation (and boundary conditions!)
4. Interpret the solution; visualize it; redesign if needed.

Numerical simulation has become a third branch of science, together with experiment and deduction. Designing the Boeing 777 was much less expensive on a computer than in a wind tunnel. Our discussion still has to move from ordinary to partial differential equations, and from linear to nonlinear.

The texts *Introduction to Applied Mathematics* and *Computational Science and Engineering* (Wellesley-Cambridge Press) develop this whole subject further—see the course page **math.mit.edu/18085** with video lectures (also on **ocw.mit.edu**). The principles remain the same, and I hope this book helps you to see the framework behind the computations.

Problem Set 8.1

1 Show that $\det A_0^{\mathrm{T}}C_0A_0 = c_1c_2c_3+c_1c_3c_4+c_1c_2c_4+c_2c_3c_4$. Find also $\det A_1^{\mathrm{T}}C_1A_1$ in the fixed-free example.

2 Invert $A_1^{\mathrm{T}}C_1A_1$ in the fixed-free example by multiplying $A_1^{-1}C_1^{-1}(A_1^{\mathrm{T}})^{-1}$.

3 In the free-free case when $A^{\mathrm{T}}CA$ in equation (9) is singular, add the three equations $A^{\mathrm{T}}CAu = f$ to show that we need $f_1 + f_2 + f_3 = 0$. Find a solution to $A^{\mathrm{T}}CAu = f$ when the forces $f = (-1, 0, 1)$ balance themselves. Find all solutions!

4 Both end conditions for the free-free differential equation are $du/dx = 0$:

$$-\frac{d}{dx}\left(c(x)\frac{du}{dx}\right) = f(x) \quad \text{with} \quad \frac{du}{dx} = 0 \text{ at both ends.}$$

Integrate both sides to show that the force $f(x)$ must balance itself, $\int f(x)\,dx = 0$, or there is no solution. The complete solution is one particular solution $u(x)$ plus any constant. The constant corresponds to $u = (1, 1, 1)$ in the nullspace of $A^{\mathrm{T}}CA$.

5 In the fixed-free problem, the matrix A is square and invertible. We can solve $A^{\mathrm{T}}y = f$ separately from $Au = e$. Do the same for the differential equation:

$$\text{Solve} \quad -\frac{dy}{dx} = f(x) \quad \text{with} \quad y(1) = 0. \quad \text{Graph} \quad y(x) \quad \text{if} \quad f(x) = 1.$$

6 The 3 by 3 matrix $K_1 = A_1^{\mathrm{T}} C_1 A_1$ in equation (6) splits into three "element matrices" $c_1 E_1 + c_2 E_2 + c_3 E_3$. Write down those pieces, one for each c. Show how they come from *column times row* multiplication of $A_1^{\mathrm{T}} C_1 A_1$. This is how finite element stiffness matrices are actually assembled.

7 For five springs and four masses with both ends fixed, what are the matrices A and C and K? With $C = I$ solve $K\boldsymbol{u} = \text{ones}(4)$.

8 Compare the solution $\boldsymbol{u} = (u_1, u_2, u_3, u_4)$ in Problem 7 to the solution of the continuous problem $-u'' = 1$ with $u(0) = 0$ and $u(1) = 0$. The parabola $u(x)$ should correspond at $x = \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}$ to $\boldsymbol{u}$—is there a $(\Delta x)^2$ factor to account for?

9 Solve the fixed-free problem $-u'' = mg$ with $u(0) = 0$ and $u'(1) = 0$. Compare $u(x)$ at $x = \frac{1}{3}, \frac{2}{3}, \frac{3}{3}$ with the vector $\boldsymbol{u} = (3mg, 5mg, 6mg)$ in Example 2.

10 Suppose $c_1 = c_2 = c_3 = c_4 = 1$, $m_1 = 2$ and $m_2 = m_3 = 1$. Solve $A^{\mathrm{T}} C A \boldsymbol{u} = (2, 1, 1)$ for this fixed-fixed line of springs. Which mass moves the most (largest u)?

11 (MATLAB) Find the displacements $u(1), \ldots, u(100)$ of 100 masses connected by springs all with $c = 1$. Each force is $f(i) = .01$. Print graphs of $\boldsymbol{u}$ with **fixed-fixed** and **fixed-free** ends. Note that diag(ones(n, 1), d) is a matrix with n ones along diagonal d. This print command will graph a vector u:

```
plot(u, '+');   xlabel('mass number');   ylabel('movement');   print
```

12 (MATLAB) Chemical engineering has a first derivative du/dx from fluid velocity as well as d^2u/dx^2 from diffusion. Replace du/dx by a *forward* difference, then a *centered* difference, then a *backward* difference, with $\Delta x = \frac{1}{8}$. Graph your three numerical solutions of

$$-\frac{d^2u}{dx^2} + 10\frac{du}{dx} = 1 \quad \text{with} \quad u(0) = u(1) = 0.$$

This ***convection-diffusion equation*** appears everywhere. It transforms to the Black-Scholes equation for option prices in mathematical finance.

Problem 12 is developed into the first MATLAB homework in my 18.085 course on Computational Science and Engineering at MIT. Videos on *ocw.mit.edu*.

8.2 Graphs and Networks

Over the years I have seen one model so often, and I found it so basic and useful, that I always put it first. The model consists of ***nodes connected by edges***. This is called a ***graph***.

Graphs of the usual kind display functions $f(x)$. Graphs of this node-edge kind lead to matrices. This section is about the ***incidence matrix*** of a graph—which tells how the n nodes are connected by the m edges. Normally $m > n$, there are more edges than nodes.

For any m by n matrix there are two fundamental subspaces in $\mathbf{R}^n$ and two in $\mathbf{R}^m$. They are the row spaces and nullspaces of A and A^{T}. Their *dimensions* are related by the most important theorem in linear algebra. The second part of that theorem is the *orthogonality* of the subspaces. Our goal is to show how examples from graphs illuminate the Fundamental Theorem of Linear Algebra.

We review the four subspaces (for any matrix). Then we construct a ***directed graph*** and its ***incidence matrix***. The dimensions will be easy to discover. But we want the subspaces themselves—this is where orthogonality helps. It is essential to connect the subspaces to the graph they come from. By specializing to incidence matrices, the laws of linear algebra become Kirchhoff's laws. Please don't be put off by the words "current" and "voltage" and "Kirchhoff." These rectangular matrices are the best.

Every entry of an incidence matrix is 0 or 1 or -1. This continues to hold during elimination. All pivots and multipliers are ± 1. Therefore both factors in $A = LU$ also contain $0, 1, -1$. So do the nullspace matrices! All four subspaces have basis vectors with these exceptionally simple components. The matrices are not concocted for a textbook, they come from a model that is absolutely essential in pure and applied mathematics.

Here is a first incidence matrix. Notice -1 and 1 in each row. This matrix takes ***differences in voltage***, across six edges of a graph. The voltages are x_1, x_2, x_3, x_4 at the four nodes in Figure 8.4—where we will construct this matrix A. Its echelon form is U:

$$\begin{array}{l} \textbf{Incidence} \\ \textbf{matrix} \\ \textbf{6 edges} \\ \textbf{4 nodes} \end{array} \quad A = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \quad \text{reduces to} \quad U = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The nullspace of A and U is the line through $\boldsymbol{x} = (1, 1, 1, 1)$. The column spaces of A and U have dimension $r = 3$. The pivot rows are a basis for the row space.

Figure 8.3 shows more—the subspaces are orthogonal. ***Every vector in the nullspace is perpendicular to every vector in the row space.*** This comes directly from the m equations $A\boldsymbol{x} = \mathbf{0}$. For A and U above, $\boldsymbol{x} = (1, 1, 1, 1)$ is perpendicular to all rows and thus to the whole row space. Equal voltages produce no current!

I would like to review the Four Fundamental Subspaces before using them. The whole point will be to see their meaning on the network.

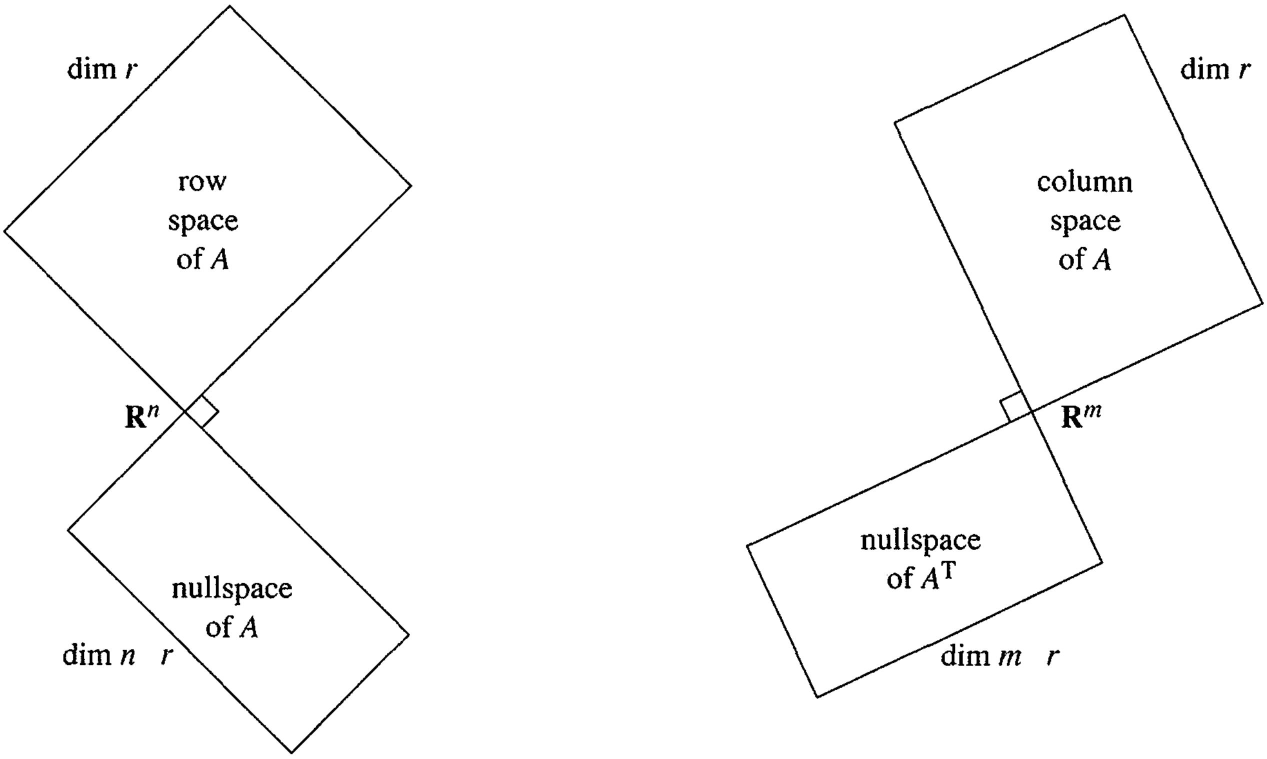


Figure 8.3: **Big picture:** The four subspaces with their dimensions and orthogonality.

Start with an m by n matrix. Its columns are vectors in $\mathbf{R}^m$. Their linear combinations produce the ***column space*** $C(A)$, a subspace of $\mathbf{R}^m$. Those combinations are exactly the matrix-vector products $A\boldsymbol{x}$.

The rows of A are vectors in $\mathbf{R}^n$ (or they would be, if they were column vectors). Their linear combinations produce the ***row space***. To avoid any inconvenience with rows, we transpose the matrix. The row space becomes $C(A^{\mathrm{T}})$, the column space of A^{T}.

The central questions of linear algebra come from these two ways of looking at the same numbers, by columns and by rows.

The ***nullspace*** $N(A)$ contains every $\boldsymbol{x}$ that satisfies $A\boldsymbol{x} = \mathbf{0}$—this is a subspace of $\mathbf{R}^n$. The ***"left" nullspace*** contains all solutions to $A^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$. Now $\boldsymbol{y}$ has m components, and $N(A^{\mathrm{T}})$ is a subspace of $\mathbf{R}^m$. Written as $\boldsymbol{y}^{\mathrm{T}}A = \mathbf{0}^{\mathrm{T}}$, we are combining rows of A to produce the zero row. The four subspaces are illustrated by Figure 8.3, which shows $\mathbf{R}^n$ on one side and $\mathbf{R}^m$ on the other. The link between them is A.

The information in that figure is crucial. First come the dimensions, which obey the two central laws of linear algebra:

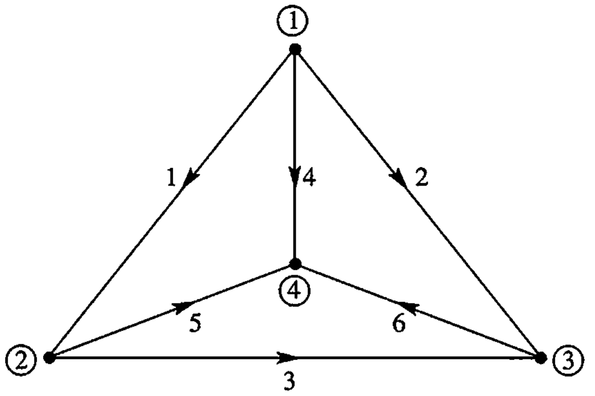
$$\dim \boldsymbol{C}(A) = \dim \boldsymbol{C}(A^{\mathrm{T}}) \quad \text{and} \quad \dim \boldsymbol{C}(A) + \dim \boldsymbol{N}(A) = \boldsymbol{n}.$$

When the row space has dimension r, the nullspace has dimension $n - r$. Elimination leaves these two spaces unchanged, and the echelon form U gives the dimension count. There are r rows and columns with pivots. There are $n - r$ free columns without pivots, and those lead to vectors in the nullspace.

This review of the subspaces applies to any matrix A—only the example was special. Now we concentrate on that example. It is the incidence matrix for a particular graph, and we look to the graph for the meaning of every subspace.

Directed Graphs and Incidence Matrices

Figure 8.4 displays a *graph* with $m = 6$ edges and $n = 4$ nodes, so the matrix A is 6 by 4. It tells which nodes are connected by which edges. The entries -1 and $+1$ also tell the direction of each arrow (this is a *directed* graph). The first row $-1, 1, 0, 0$ of A gives a record of the first edge from node 1 to node 2:



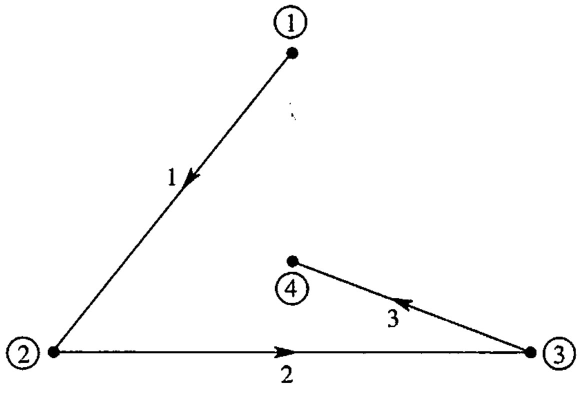
$$A=\begin{array}{cccc|cl}
\multicolumn{4}{c|}{\textbf{node}} & & \\
\textcircled{1} & \textcircled{2} & \textcircled{3} & \textcircled{4} & & \\
\hline
-1 & 1 & 0 & 0 & 1 & \\
-1 & 0 & 1 & 0 & 2 & \\
0 & -1 & 1 & 0 & 3 & \\
-1 & 0 & 0 & 1 & 4 & \textbf{edge} \\
0 & -1 & 0 & 1 & 5 & \\
0 & 0 & -1 & 1 & 6 &
\end{array}$$

Figure 8.4a: Complete graph with $m = 6$ edges and $n = 4$ nodes.

Row numbers are edge numbers, column numbers are node numbers.

You can write down A immediately by looking at the graph.

The second graph has the same four nodes but only three edges. Its incidence matrix is 3 by 4:



$$B=\begin{array}{cccc|cl}
\multicolumn{4}{c|}{\textbf{node}} & & \\
\textcircled{1} & \textcircled{2} & \textcircled{3} & \textcircled{4} & & \\
\hline
1 & 1 & 0 & 0 & 1 & \\
0 & 1 & 1 & 0 & 2 & \textbf{edge} \\
0 & 0 & 1 & 1 & 3 &
\end{array}$$

Figure 8.4b: Tree with 3 edges and 4 nodes and no loops.

The first graph is *complete*—every pair of nodes is connected by an edge. The second graph is a ***tree***—the graph has ***no closed loops***. Those graphs are the two extremes, the maximum number of edges is $\frac{1}{2}n(n-1)$ and the minimum (a tree) is $m = n - 1$.

The rows of B match the nonzero rows of U—the echelon form found earlier. ***Elimination reduces every graph to a tree.*** The loops produce zero rows in U. Look at the loop from edges 1, 2, 3 in the first graph, which leads to a zero row:

$$\begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 1 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

Those steps are typical. When two edges share a node, elimination produces the "shortcut edge" without that node. If the graph already has this shortcut edge, elimination gives a row of zeros. When the dust clears we have a tree.

An idea suggests itself: ***Rows are dependent when edges form a loop.*** Independent rows come from trees. This is the key to the row space. We are assuming that the graph is connected, and it makes no fundamental difference which way the arrows go. On each edge, flow with the arrow is "positive." Flow in the opposite direction counts as negative. The flow might be a current or a signal or a force—or even oil or gas or water.

For the column space we look at Ax, which is a vector of differences:

$$Ax = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_2 - x_1 \\ x_3 - x_1 \\ x_3 - x_2 \\ x_4 - x_1 \\ x_4 - x_2 \\ x_4 - x_3 \end{bmatrix}. \tag{1}$$

The unknowns x_1, x_2, x_3, x_4 represent ***potentials*** or ***voltages*** at the nodes. Then Ax gives the ***potential differences*** or ***voltage differences*** across the edges. It is these differences that cause flows. We now examine the meaning of each subspace.

1 The ***nullspace*** contains the solutions to $Ax = \mathbf{0}$. All six potential differences are zero. This means: *All four potentials are equal.* Every x in the nullspace is a constant vector (c, c, c, c). The nullspace of A is a line in $\mathbf{R}^n$—its dimension is $n - r = 1$.

The second incidence matrix B has the same nullspace. It contains $(\mathbf{1}, \mathbf{1}, \mathbf{1}, \mathbf{1})$:

$$Bx = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

We can raise or lower all potentials by the same amount c, without changing the differences. There is an "arbitrary constant" in the potentials. Compare this with the same statement for functions. We can raise or lower $f(x)$ by the same amount C, without changing its derivative. There is an arbitrary constant C in the integral.

Calculus adds "$+C$" to indefinite integrals. Graph theory adds (c, c, c, c) to the vector x of potentials. Linear algebra adds any vector x_n in the nullspace to one particular solution of $Ax = b$.

The "$+C$" disappears in calculus when the integral starts at a known point $x = a$. Similarly the nullspace disappears when we set $x_4 = 0$. The unknown x_4 is removed and so are the fourth columns of A and B. Electrical engineers would say that node 4 has been "grounded."

2 The ***row space*** contains all combinations of the six rows. Its dimension is certainly not six. The equation $r + (n - r) = n$ must be $3 + 1 = 4$. The rank is $r = 3$, as we also saw from elimination. After 3 edges, we start forming loops! The new rows are not independent.

How can we tell if $v = (v_1, v_2, v_3, v_4)$ is in the row space? The slow way is to combine rows. The quick way is by orthogonality:

v is in the row space if and only if it is perpendicular to $(1, 1, 1, 1)$ in the nullspace.

The vector $v = (0, 1, 2, 3)$ fails this test—its components add to 6. The vector $(-6, 1, 2, 3)$ passes the test. It lies in the row space because its components add to zero. It equals 6(row 1) + 5(row 3) + 3(row 6).

Each row of A adds to zero. This must be true for every vector in the row space.

3 The ***column space*** contains all combinations of the four columns. We expect three independent columns, since there were three independent rows. The first three columns are independent (so are any three). But the four columns add to the zero vector, which says again that $(1, 1, 1, 1)$ is in the nullspace. ***How can we tell if a particular vector b is in the column space of an incidence matrix?***

First answer Try to solve $Ax = b$. That misses all the insight. As before, orthogonality gives a better answer. We are now coming to Kirchhoff's two famous laws of circuit theory—the voltage law and current law. Those are natural expressions of "laws" of linear algebra. It is especially pleasant to see the key role of the left nullspace.

Second answer Ax is the vector of differences in equation (1). If we add differences around a closed loop in the graph, the cancellation leaves zero. Around the big triangle formed by edges $1, 3, -2$ (the arrow goes backward on edge 2) the differences cancel:

$$\textbf{Voltage Law} \qquad (x_2 - x_1) + (x_3 - x_2) - (x_3 - x_1) = 0.$$

The components of Ax add to zero around every loop. When b is in the column space of A, it must obey the same law:

$$\textbf{\textit{Kirchhoff's Law:}} \qquad b_1 + b_3 - b_2 = 0.$$

By testing each loop, we decide whether b is in the column space. $Ax = b$ can be solved exactly when the components of b satisfy all the same dependencies as the rows of A. Then elimination leads to $0 = 0$, and $Ax = b$ is consistent.

4 The ***left nullspace*** contains the solutions to $A^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$. Its dimension is $m - r = 6 - 3$:

Current Law (KCL)

$$A^{\mathrm{T}}y = \begin{bmatrix} -1 & -1 & 0 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \qquad (2)$$

The true number of equations is $r = 3$ and not $n = 4$. Reason: The four equations add to $0 = 0$. The fourth equation follows automatically from the first three.

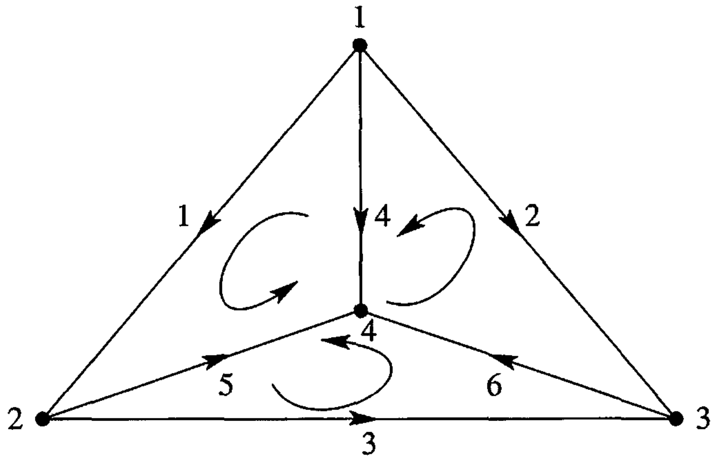
What do the equations mean? The first equation says that $-y_1 - y_2 - y_4 = 0$. ***The net flow into node* 1 *is zero.*** The fourth equation says that $y_4 + y_5 + y_6 = 0$. ***Flow into the node minus flow out is zero.*** The equations $A^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$ are famous and fundamental:

Kirchhoff's Current Law: ***Flow in equals flow out at each node.***

This law deserves first place among the equations of applied mathematics. It expresses "*conservation*" and "*continuity*" and "*balance*." Nothing is lost, nothing is gained. When currents or forces are in equilibrium, the equation to solve is $A^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$. Notice the beautiful fact that the matrix in this balance equation is the transpose of the incidence matrix A.

What are the actual solutions to $A^{\mathrm{T}}y = \mathbf{0}$? The currents must balance themselves. The easiest way is to **flow around a loop**. If a unit of current goes around the big triangle (forward on edge 1, forward on 3, backward on 2), the vector is $y = (1, -1, 1, 0, 0, 0)$. This satisfies $A^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$. *Every loop current is a solution to the Current Law.* Around the loop, flow in equals flow out at every node. A smaller loop goes forward on edge 1, forward on 5, back on 4. Then $y = (1, 0, 0, -1, 1, 0)$ is also in the left nullspace.

We expect three independent y's, since $6 - 3 = 3$. The three small loops in the graph are independent. The big triangle seems to give a fourth y, but it is the sum of flows around the small loops. The small loops give a basis for the left nullspace.



$$\underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}}_{\text{small loops}} = \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{\text{big loop}}$$

Summary The incidence matrix A comes from a connected graph with n nodes and m edges. The row space and column space have dimensions $n-1$. The nullspaces of A and A^{T} have dimension 1 and $m-n+1$:

1 The constant vectors $(c, c, \ldots, c)$ make up the nullspace of A.

2 There are $r = n - 1$ independent rows, using edges from any tree.

3 ***Voltage law***: The components of Ax add to zero around every loop.

4 ***Current law***: $A^{\mathrm{T}}y = \mathbf{0}$ is solved by loop currents. $N(A^{\mathrm{T}})$ has dimension $m - r$. ***There are $m - r = m - n + 1$ independent loops in the graph.***

For every graph in a plane, linear algebra yields ***Euler's formula***:

$$(\textit{number of nodes}) - (\textit{number of edges}) + (\textit{number of small loops}) = \mathbf{1}.$$

This is $\boldsymbol{n - m + (m - n + 1) = 1}$. The graph in our example has $4 - 6 + 3 = 1$.

A single triangle has (3 nodes) − (3 edges) + (1 loop). On a 10-node tree with 9 edges and no loops, Euler's count is $10 - 9 + 0$. All planar graphs lead to the answer 1.

Networks and $A^{\mathrm{T}}CA$

In a real network, the current y along an edge is the product of two numbers. One number is the difference between the potentials x at the ends of the edge. This difference is Ax and it drives the flow. The other number is the "***conductance***" c—which measures how easily flow gets through.

In physics and engineering, c is decided by the material. For electrical currents, c is high for metal and low for plastics. For a superconductor, c is nearly infinite. If we consider elastic stretching, c might be low for metal and higher for plastics. In economics, c measures the capacity of an edge or its cost.

To summarize, the graph is known from its "connectivity matrix" A. This tells the connections between nodes and edges. A ***network*** goes further, and assigns a conductance c to each edge. These numbers $c_1, \ldots, c_m$ go into the "conductance matrix" C—which is diagonal.

For a network of resistors, the conductance is $c = 1/(\text{resistance})$. In addition to Kirchhoff's Laws for the whole system of currents, we have Ohm's Law for each particular current. Ohm's Law connects the current y_1 on edge 1 to the potential difference $x_2 - x_1$ between the nodes:

Ohm's Law: Current along edge = conductance times potential difference.

Ohm's Law for all m currents is $y = -CAx$. The vector Ax gives the potential differences, and C multiplies by the conductances. Combining Ohm's Law with Kirchhoff's Current

Law $A^{\mathrm{T}}y = \mathbf{0}$, we get $A^{\mathrm{T}}CAx = \mathbf{0}$. This is *almost* the central equation for network flows. The only thing wrong is the zero on the right side! The network needs power from outside—a voltage source or a current source—to make something happen.

Note about signs In circuit theory we change from Ax to $-Ax$. The flow is from higher potential to lower potential. There is (positive) current from node 1 to node 2 when $x_1 - x_2$ is positive—whereas Ax was constructed to yield $x_2 - x_1$. The minus sign in physics and electrical engineering is a plus sign in mechanical engineering and economics. Ax versus $-Ax$ is a general headache but unavoidable.

Note about applied mathematics Every new application has its own form of Ohm's law. For elastic structures $y = CAx$ is Hooke's law. The stress y is (elasticity C) times (stretching Ax). For heat conduction, Ax is a temperature gradient. For oil flows it is a pressure gradient. There is a similar law in Section 8.6 for least squares regression in statistics.

My textbooks *Introduction to Applied Mathematics* and *Computational Science and Engineering* (Wellesley-Cambridge Press) are practically built on $A^{\mathrm{T}}CA$. This is the key to equilibrium in matrix equations and also in differential equations. Applied mathematics is more organized than it looks. *I have learned to watch for $A^{\mathrm{T}}CA$.*

We now give an example with a current source. Kirchhoff's Law changes from $A^{\mathrm{T}}y = \mathbf{0}$ to $A^{\mathrm{T}}y = f$, to balance the source f from outside. *Flow into each node still equals flow out.* Figure 8.5 shows the network with its conductances $c_1, \ldots, c_6$, and it shows the current source going into node 1. The source comes out at node 4 to keep the balance (in = out). The problem is: ***Find the currents $y_1, \ldots, y_6$ on the six edges.***

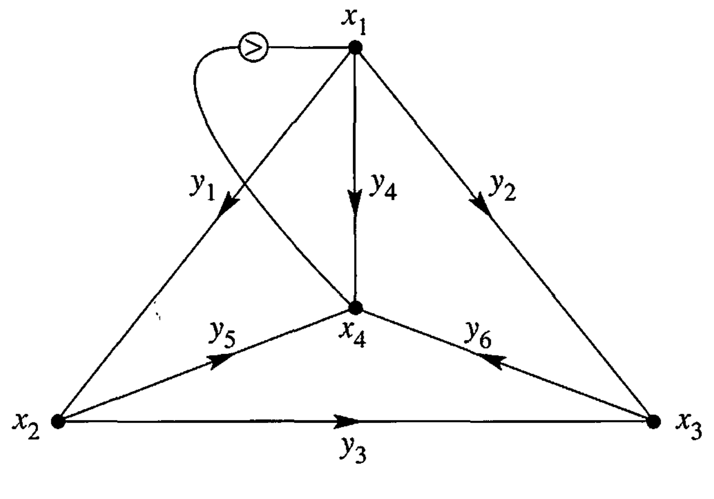


Figure 8.5: The currents in a network with a source S into node 1.

Example 1 All conductances are $c = 1$, so that $C = I$. A current y_4 travels directly from node 1 to node 4. Other current goes the long way from node 1 to node 2 to node 4 (this is $y_1 = y_5$). Current also goes from node 1 to node 3 to node 4 (this is $y_2 = y_6$). We can find the six currents by using special rules for symmetry, or we can do it right by using

$A^{\mathrm{T}}CA$. Since $C = I$, this matrix is $A^{\mathrm{T}}A$, the ***graph Laplacian matrix***:

$$\begin{bmatrix} -1 & -1 & 0 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix}$$

That last matrix is not invertible! We cannot solve for all four potentials because (1, 1, 1, 1) is in the nullspace. One node has to be grounded. Setting $x_4 = 0$ removes the fourth row and column, and this leaves a 3 by 3 invertible matrix. Now we solve $A^{\mathrm{T}}CAx = f$ for the unknown potentials x_1, x_2, x_3, with source S into node 1:

$$\textbf{Voltages} \qquad \begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} S \\ 0 \\ 0 \end{bmatrix} \quad \text{gives} \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} S/2 \\ S/4 \\ S/4 \end{bmatrix}.$$

Ohm's Law $y = -CAx$ yields the six currents. Remember $C = I$ and $x_4 = 0$:

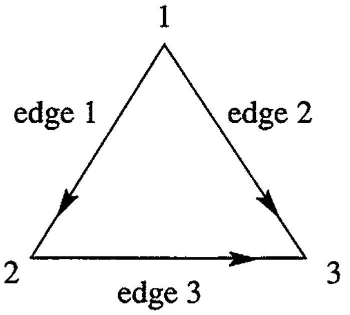
$$\textbf{Currents} \qquad \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} = - \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} S/2 \\ S/4 \\ S/4 \\ 0 \end{bmatrix} = \begin{bmatrix} S/4 \\ S/4 \\ 0 \\ S/2 \\ S/4 \\ S/4 \end{bmatrix}.$$

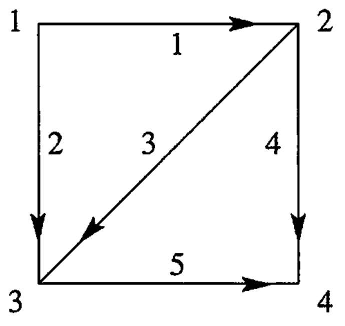
Half the current goes directly on edge 4. That is $y_4 = S/2$. No current crosses from node 2 to node 3. Symmetry indicated $y_3 = 0$ and now the solution proves it.

The same matrix $A^{\mathrm{T}}A$ appears in least squares. Nature distributes the currents to minimize the heat loss. Statistics chooses $\widehat{x}$ to minimize the least squares error.

Problem Set 8.2

Problems 1–7 and 8–14 are about the incidence matrices for these graphs.





1 Write down the 3 by 3 incidence matrix A for the triangle graph. The first row has -1 in column 1 and $+1$ in column 2. What vectors (x_1, x_2, x_3) are in its nullspace? How do you know that $(1,0,0)$ is not in its row space?

2 Write down A^{T} for the triangle graph. Find a vector $\boldsymbol{y}$ in its nullspace. The components of $\boldsymbol{y}$ are currents on the edges—how much current is going around the triangle?

3 Eliminate x_1 and x_2 from the third equation to find the echelon matrix U. What tree corresponds to the two nonzero rows of U?

$$\begin{aligned} -x_1 + x_2 &= b_1 \\ -x_1 + x_3 &= b_2 \\ -x_2 + x_3 &= b_3. \end{aligned}$$

4 Choose a vector (b_1, b_2, b_3) for which $A\boldsymbol{x} = \boldsymbol{b}$ can be solved, and another vector $\boldsymbol{b}$ that allows no solution. How are those $\boldsymbol{b}$'s related to $\boldsymbol{y} = (1,-1,1)$?

5 Choose a vector (f_1, f_2, f_3) for which $A^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{f}$ can be solved, and a vector $\boldsymbol{f}$ that allows no solution. How are those $\boldsymbol{f}$'s related to $\boldsymbol{x} = (1,1,1)$? The equation $A^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{f}$ is Kirchhoff's _____ law.

6 Multiply matrices to find $A^{\mathrm{T}}A$. Choose a vector $\boldsymbol{f}$ for which $A^{\mathrm{T}}A\boldsymbol{x} = \boldsymbol{f}$ can be solved, and solve for $\boldsymbol{x}$. Put those potentials $\boldsymbol{x}$ and the currents $\boldsymbol{y} = -A\boldsymbol{x}$ and current sources $\boldsymbol{f}$ onto the triangle graph. Conductances are 1 because $C = I$.

7 With conductances $c_1 = 1$ and $c_2 = c_3 = 2$, multiply matrices to find $A^{\mathrm{T}}CA$. For $\boldsymbol{f} = (1,0,-1)$ find a solution to $A^{\mathrm{T}}CA\boldsymbol{x} = \boldsymbol{f}$. Write the potentials $\boldsymbol{x}$ and currents $\boldsymbol{y} = -CA\boldsymbol{x}$ on the triangle graph, when the current source $\boldsymbol{f}$ goes into node 1 and out from node 3.

8 Write down the 5 by 4 incidence matrix A for the square graph with two loops. Find one solution to $A\boldsymbol{x} = \boldsymbol{0}$ and two solutions to $A^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{0}$.

9 Find two requirements on the b's for the five differences $x_2 - x_1, x_3 - x_1, x_3 - x_2, x_4 - x_2, x_4 - x_3$ to equal b_1, b_2, b_3, b_4, b_5. You have found Kirchhoff's _____ law around the two _____ in the graph.

10 Reduce A to its echelon form U. The three nonzero rows give the incidence matrix for what graph? You found one tree in the square graph—find the other seven trees.

11 Multiply matrices to find $A^{\mathrm{T}}A$ and guess how its entries come from the graph:

(a) The diagonal of $A^{\mathrm{T}}A$ tells how many _____ into each node.

(b) The off-diagonals -1 or 0 tell which pairs of nodes are _____.

12 Why is each statement true about $A^{\mathrm{T}}A$? *Answer for $A^{\mathrm{T}}A$ not A.*

(a) Its nullspace contains $(1,1,1,1)$. Its rank is $n-1$.

(b) It is positive semidefinite but not positive definite.

(c) Its four eigenvalues are real and their signs are _____.

13 With conductances $c_1 = c_2 = 2$ and $c_3 = c_4 = c_5 = 3$, multiply the matrices $A^{\mathrm{T}}CA$. Find a solution to $A^{\mathrm{T}}CA\boldsymbol{x} = \boldsymbol{f} = (1,0,0,-1)$. Write these potentials $\boldsymbol{x}$ and currents $\boldsymbol{y} = -CA\boldsymbol{x}$ on the nodes and edges of the square graph.

14 The matrix $A^{\mathrm{T}}CA$ is not invertible. What vectors $\boldsymbol{x}$ are in its nullspace? Why does $A^{\mathrm{T}}CA\boldsymbol{x} = \boldsymbol{f}$ have a solution if and only if $f_1 + f_2 + f_3 + f_4 = 0$?

15 A connected graph with 7 nodes and 7 edges has how many loops?

16 For the graph with 4 nodes, 6 edges, and 3 loops, add a new node. If you connect it to one old node, Euler's formula becomes $(\quad) - (\quad) + (\quad) = 1$. If you connect it to two old nodes, Euler's formula becomes $(\quad) - (\quad) + (\quad) = 1$.

17 Suppose A is a 12 by 9 incidence matrix from a connected (but unknown) graph.

(a) How many columns of A are independent?

(b) What condition on $\boldsymbol{f}$ makes it possible to solve $A^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{f}$?

(c) The diagonal entries of $A^{\mathrm{T}}A$ give the number of edges into each node. What is the sum of those diagonal entries?

18 Why does a complete graph with $n = 6$ nodes have $m = 15$ edges? A tree connecting 6 nodes has _____ edges.

Note The ***stoichiometric matrix*** in chemistry is an important "generalized" incidence matrix. Its entries show how much of each chemical species (each column) goes into each reaction (each row).

8.3 Markov Matrices, Population, and Economics

This section is about ***positive matrices***: every $a_{ij} > 0$. The key fact is quick to state: ***The largest eigenvalue is real and positive and so is its eigenvector.*** In economics and ecology and population dynamics and random walks, that fact leads a long way:

$$\textbf{Markov}\quad \lambda_{\max} = 1 \qquad \textbf{Population}\quad \lambda_{\max} > 1 \qquad \textbf{Consumption}\quad \lambda_{\max} < 1$$

$\lambda_{\max}$ controls the powers of A. We will see this first for $\lambda_{\max} = 1$.

Markov Matrices

Suppose we multiply a positive vector $\boldsymbol{u}_0 = (a, 1 - a)$ again and again by this A:

$$\textbf{Markov matrix}\qquad A = \begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix} \qquad \boldsymbol{u}_1 = A\boldsymbol{u}_0 \qquad \boldsymbol{u}_2 = A\boldsymbol{u}_1 = A^2\boldsymbol{u}_0$$

After k steps we have $A^k\boldsymbol{u}_0$. The vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \ldots$ will approach a "*steady state*" $\boldsymbol{u}_\infty = (.6, .4)$. This final outcome does not depend on the starting vector: ***For every $\boldsymbol{u}_0$ we converge to the same $\boldsymbol{u}_\infty$***. The question is why.

The steady state equation $A\boldsymbol{u}_\infty = \boldsymbol{u}_\infty$ makes $\boldsymbol{u}_\infty$ ***an eigenvector with eigenvalue*** **1**:

$$\textbf{Steady state}\qquad \begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix} \begin{bmatrix} .6 \\ .4 \end{bmatrix} = \begin{bmatrix} .6 \\ .4 \end{bmatrix}.$$

Multiplying by A does not change $\boldsymbol{u}_\infty$. But this does not explain why all vectors $\boldsymbol{u}_0$ lead to $\boldsymbol{u}_\infty$. Other examples might have a steady state, but it is not necessarily attractive:

$$\textbf{Not Markov}\quad B = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \quad \text{has the unattractive steady state} \quad B \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

In this case, the starting vector $\boldsymbol{u}_0 = (0, 1)$ will give $\boldsymbol{u}_1 = (0, 2)$ and $\boldsymbol{u}_2 = (0, 4)$. The second components are doubled. In the language of eigenvalues, B has $\lambda = 1$ but also $\lambda = 2$— this produces instability. The component of $\boldsymbol{u}$ along that unstable eigenvector is multiplied by λ, and $|\lambda| > 1$ means blowup.

This section is about two special properties of A that guarantee a stable steady state. These properties define a ***Markov matrix***, and A above is one particular example:

Markov matrix

1. ***Every entry of A is nonnegative.***
2. ***Every column of A adds to* 1.**

B did not have Property 2. When A is a Markov matrix, two facts are immediate:

1. Multiplying a nonnegative $\boldsymbol{u}_0$ by A produces a nonnegative $\boldsymbol{u}_1 = A\boldsymbol{u}_0$.

2. If the components of $\boldsymbol{u}_0$ add to 1, so do the components of $\boldsymbol{u}_1 = A\boldsymbol{u}_0$.

Reason: The components of $\boldsymbol{u}_0$ add to 1 when $\begin{bmatrix}1 & \cdots & 1\end{bmatrix}\boldsymbol{u}_0 = 1$. This is true for each column of A by Property 2. Then by matrix multiplication $[1 \; \ldots \; 1]A = [1 \; \ldots \; 1]$:

$$\textbf{Components of } A\boldsymbol{u}_0 \textbf{ add to 1} \qquad [1 \;\cdots\; 1]A\boldsymbol{u}_0 = [1 \;\cdots\; 1]\boldsymbol{u}_0 = 1.$$

The same facts apply to $\boldsymbol{u}_2 = A\boldsymbol{u}_1$ and $\boldsymbol{u}_3 = A\boldsymbol{u}_2$. ***Every vector $A^k\boldsymbol{u}_0$ is nonnegative with components adding to*** 1. These are "***probability vectors.***" The limit $\boldsymbol{u}_\infty$ is also a probability vector—but we have to prove that there is a limit. We will show that $\lambda_{\max} = 1$ for a positive Markov matrix.

Example 1 The fraction of rental cars in Denver starts at $\frac{1}{50} = .02$. The fraction outside Denver is .98. Every month, 80% of the Denver cars stay in Denver (and 20% leave). Also 5% of the outside cars come in (95% stay outside). This means that the fractions $\boldsymbol{u}_0 = (.02, .98)$ are multiplied by A:

$$\textbf{First month} \qquad A = \begin{bmatrix} .80 & .05 \\ .20 & .95 \end{bmatrix} \quad \text{leads to} \quad \boldsymbol{u}_1 = A\boldsymbol{u}_0 = A\begin{bmatrix} .02 \\ .98 \end{bmatrix} = \begin{bmatrix} .065 \\ .935 \end{bmatrix}.$$

Notice that $.065 + .935 = 1$. All cars are accounted for. Each step multiplies by A:

$$\textbf{Next month} \qquad \boldsymbol{u}_2 = A\boldsymbol{u}_1 = (.09875,\ .90125). \text{ This is } A^2\boldsymbol{u}_0.$$

All these vectors are positive because A is positive. Each vector $\boldsymbol{u}_k$ will have its components adding to 1. The first component has grown from .02 and cars are moving toward Denver. What happens in the long run?

This section involves powers of matrices. The understanding of A^k was our first and best application of diagonalization. Where A^k can be complicated, the diagonal matrix Λ^k is simple. The eigenvector matrix S connects them: A^k equals $S\Lambda^kS^{-1}$. The new application to Markov matrices uses the eigenvalues (in Λ) and the eigenvectors (in S). We will show that $\boldsymbol{u}_\infty$ is an eigenvector corresponding to $\lambda = 1$.

Since every column of A adds to 1, nothing is lost or gained. We are moving rental cars or populations, and no cars or people suddenly appear (or disappear). The fractions add to 1 and the matrix A keeps them that way. The question is how they are distributed after k time periods—which leads us to A^k.

Solution $A^k\boldsymbol{u}_0$ gives the fractions in and out of Denver after k steps. We diagonalize A to understand A^k. The eigenvalues are $\lambda = 1$ and .75 (the trace is 1.75).

$$Ax = \lambda x \qquad A\begin{bmatrix} .2 \\ .8 \end{bmatrix} = 1\begin{bmatrix} .2 \\ .8 \end{bmatrix} \quad \textbf{and} \quad A\begin{bmatrix} -1 \\ 1 \end{bmatrix} = .75\begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

The starting vector $\boldsymbol{u}_0$ combines $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, in this case with coefficients 1 and .18:

$$\textbf{Combination of eigenvectors} \qquad \boldsymbol{u}_0 = \begin{bmatrix} .02 \\ .98 \end{bmatrix} = \begin{bmatrix} .2 \\ .8 \end{bmatrix} + .18\begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Now multiply by A to find $\boldsymbol{u}_1$. The eigenvectors are multiplied by $\lambda_1 = 1$ and $\lambda_2 = .75$:

$$\textbf{Each } \boldsymbol{x} \textbf{ is multiplied by } \boldsymbol{\lambda} \qquad \boldsymbol{u}_1 = 1\begin{bmatrix} .2 \\ .8 \end{bmatrix} + (.75)(.18)\begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Every month, another .75 multiplies the vector x_2. The eigenvector x_1 is unchanged:

$$\textbf{After } k \textbf{ steps} \qquad u_k = A^k u_0 = \begin{bmatrix} .2 \\ .8 \end{bmatrix} + (.75)^k(.18)\begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

This equation reveals what happens. ***The eigenvector*** x_1 ***with*** $\lambda = 1$ ***is the steady state.*** The other eigenvector x_2 disappears because $|\lambda| < 1$. The more steps we take, the closer we come to $u_\infty = (.2, .8)$. In the limit, $\frac{2}{10}$ of the cars are in Denver and $\frac{8}{10}$ are outside. This is the pattern for Markov chains, even starting from $u_0 = (0, 1)$:

If A is a *positive* Markov matrix (entries $a_{ij} > 0$, each column adds to **1**), then $\lambda_1 = 1$ is larger than any other eigenvalue. The eigenvector x_1 is the ***steady state***:

$$u_k = x_1 + c_2(\lambda_2)^k x_2 + \cdots + c_n(\lambda_n)^k x_n \quad \textit{always approaches} \quad u_\infty = x_1.$$

The first point is to see that $\lambda = 1$ is an eigenvalue of A. *Reason*: Every column of $A - I$ adds to $1 - 1 = 0$. The rows of $A - I$ add up to the zero row. Those rows are linearly dependent, so $A - I$ is singular. Its determinant is zero and $\lambda = 1$ is an eigenvalue.

The second point is that no eigenvalue can have $|\lambda| > 1$. With such an eigenvalue, the powers A^k would grow. But A^k is also a Markov matrix! A^k has nonnegative entries still adding to 1—and that leaves no room to get large.

A lot of attention is paid to the possibility that another eigenvalue has $|\lambda| = 1$.

Example 2 $A = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ has no steady state because $\lambda_2 = -1$.

This matrix sends all cars from inside Denver to outside, and vice versa. The powers A^k alternate between A and I. The second eigenvector $x_2 = (-1, 1)$ will be multiplied by $\lambda_2 = -1$ at every step—and does not become smaller: No steady state.

Suppose the entries of A or any power of A are all *positive*—zero is not allowed. In this "regular" or "primitive" case, $\lambda = 1$ is strictly larger than any other eigenvalue. The powers A^k approach the rank one matrix that has the steady state in every column.

Example 3 (**"Everybody moves"**) Start with three groups. At each time step, half of group 1 goes to group 2 and the other half goes to group 3. The other groups also *split in half and move*. Take one step from the starting populations p_1, p_2, p_3:

$$\textbf{New populations} \qquad u_1 = Au_0 = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}p_2 + \frac{1}{2}p_3 \\ \frac{1}{2}p_1 + \frac{1}{2}p_3 \\ \frac{1}{2}p_1 + \frac{1}{2}p_2 \end{bmatrix}.$$

A is a Markov matrix. Nobody is born or lost. A contains zeros, which gave trouble in Example 2. But after two steps in this new example, the zeros disappear from A^2:

$$\textbf{Two-step matrix} \qquad u_2 = A^2 u_0 = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}.$$

The eigenvalues of A are $\lambda_1 = 1$ (because A is Markov) and $\lambda_2 = \lambda_3 = -\frac{1}{2}$. For $\lambda = 1$, ***the eigenvector*** $x_1 = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ ***will be the steady state.*** When three equal populations split in half and move, the populations are again equal. Starting from $u_0 = (8, 16, 32)$, the Markov chain approaches its steady state:

$$u_0 = \begin{bmatrix} 8 \\ 16 \\ 32 \end{bmatrix} \qquad u_1 = \begin{bmatrix} 24 \\ 20 \\ 12 \end{bmatrix} \qquad u_2 = \begin{bmatrix} 16 \\ 18 \\ 22 \end{bmatrix} \qquad u_3 = \begin{bmatrix} 20 \\ 19 \\ 17 \end{bmatrix}.$$

The step to u_4 will split some people in half. This cannot be helped. The total population is $8 + 16 + 32 = 56$ at every step. The steady state is 56 times $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. You can see the three populations approaching, but never reaching, their final limits $56/3$.

Challenge Problem 6.7.16 created a Markov matrix A from the number of links between websites. The steady state u will give the Google rankings. ***Google finds*** u_∞ ***by a random walk that follows links*** (*random surfing*). That eigenvector comes from counting the fraction of visits to each website—a quick way to compute the steady state.

The size $|\lambda_2|$ of the next largest eigenvalue controls the speed of convergence to steady state.

Perron-Frobenius Theorem

One matrix theorem dominates this subject. The Perron-Frobenius Theorem applies when all $a_{ij} \geq 0$. There is no requirement that columns add to 1. We prove the neatest form, when all $a_{ij} > 0$.

Perron-Frobenius for $A > 0$ ***All numbers in $Ax = \lambda_{\max} x$ are strictly positive.***

Proof The key idea is to look at all numbers t such that $Ax \geq tx$ for some nonnegative vector x (other than $x = 0$). We are allowing inequality in $Ax \geq tx$ in order to have many positive candidates t. For the largest value $t_{\max}$ (which is attained), we will show that ***equality holds***: $Ax = t_{\max} x$.

Otherwise, if $Ax \geq t_{\max} x$ is not an equality, multiply by A. Because A is positive that produces a strict inequality $A^2 x > t_{\max} Ax$. Therefore the positive vector $y = Ax$ satisfies $Ay > t_{\max} y$, and $t_{\max}$ could be increased. This contradiction forces the equality $Ax = t_{\max} x$, and we have an eigenvalue. Its eigenvector x is positive because on the left side of that equality, Ax is sure to be positive.

To see that no eigenvalue can be larger than $t_{\max}$, suppose $Az = \lambda z$. Since λ and z may involve negative or complex numbers, we take absolute values: $|\lambda||z| = |Az| \leq A|z|$ by the "triangle inequality." This $|z|$ is a nonnegative vector, so $|\lambda|$ is one of the possible candidates t. Therefore $|\lambda|$ cannot exceed $t_{\max}$—which must be $\lambda_{\max}$.

Population Growth

Divide the population into three age groups: age < 20, age 20 to 39, and age 40 to 59. At year T the sizes of those groups are n_1, n_2, n_3. Twenty years later, the sizes have changed for two reasons:

1. **Reproduction** $n_1^{\text{new}} = F_1 n_1 + F_2 n_2 + F_3 n_3$ gives a new generation

2. **Survival** $n_2^{\text{new}} = P_1 n_1$ and $n_3^{\text{new}} = P_2 n_2$ gives the older generations

The fertility rates are F_1, F_2, F_3 (F_2 largest). The *Leslie matrix* A might look like this:

$$\begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix}^{\textbf{new}} = \begin{bmatrix} F_1 & F_2 & F_3 \\ P_1 & 0 & 0 \\ 0 & P_2 & 0 \end{bmatrix} \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} = \begin{bmatrix} .04 & \mathbf{1.1} & .01 \\ .98 & 0 & 0 \\ 0 & \mathbf{.92} & 0 \end{bmatrix} \begin{bmatrix} n_1 \\ \boldsymbol{n_2} \\ n_3 \end{bmatrix}.$$

This is population projection in its simplest form, the same matrix A at every step. In a realistic model, A will change with time (from the environment or internal factors). Professors may want to include a fourth group, age ≥ 60, but we don't allow it.

The matrix has $A \geq 0$ but not $A > 0$. The Perron-Frobenius theorem still applies because $A^3 > 0$. The largest eigenvalue is $\lambda_{\max} \approx 1.06$. You can watch the generations move, starting from $n_2 = 1$ in the middle generation:

$$\textbf{eig}(A) = \begin{matrix} \mathbf{1.06} \\ -1.01 \\ -0.01 \end{matrix} \quad A^2 = \begin{bmatrix} 1.08 & \mathbf{0.05} & .00 \\ 0.04 & \mathbf{1.08} & .01 \\ 0.90 & 0 & 0 \end{bmatrix} \quad A^3 = \begin{bmatrix} 0.10 & \mathbf{1.19} & .01 \\ 0.06 & \mathbf{0.05} & .00 \\ 0.04 & \mathbf{0.99} & .01 \end{bmatrix}.$$

A fast start would come from $\boldsymbol{u}_0 = (0, 1, 0)$. That middle group will reproduce 1.1 and also survive .92. The newest and oldest generations are in $\boldsymbol{u}_1 = (1.1, 0, .92) =$ column 2 of A. Then $\boldsymbol{u}_2 = A\boldsymbol{u}_1 = A^2\boldsymbol{u}_0$ is the second column of A^2. The early numbers (transients) depend a lot on $\boldsymbol{u}_0$, but ***the asymptotic growth rate*** $\lambda_{\max}$ ***is the same from every start.*** Its eigenvector $\boldsymbol{x} = (.63, .58, .51)$ shows all three groups growing steadily together.

Caswell's book on *Matrix Population Models* emphasizes sensitivity analysis. The model is never exactly right. If the F's or P's in the matrix change by 10%, does $\lambda_{\max}$ go below 1 (which means extinction)? Problem 19 will show that a matrix change ΔA produces an eigenvalue change $\Delta\lambda = \boldsymbol{y}^{\mathrm{T}}(\Delta A)\boldsymbol{x}$. Here $\boldsymbol{x}$ and $\boldsymbol{y}^{\mathrm{T}}$ are the right and left eigenvectors of A. So $\boldsymbol{x}$ is a column of S and $\boldsymbol{y}^{\mathrm{T}}$ is a row of S^{-1}.

Linear Algebra in Economics: The Consumption Matrix

A long essay about linear algebra in economics would be out of place here. A short note about one matrix seems reasonable. The ***consumption matrix*** tells how much of each input goes into a unit of output. This describes the manufacturing side of the economy.

Consumption matrix We have n industries like chemicals, food, and oil. To produce a unit of chemicals may require .2 units of chemicals, .3 units of food, and .4 units of oil. Those numbers go into row 1 of the consumption matrix A:

$$\begin{bmatrix} \text{chemical output} \\ \text{food output} \\ \text{oil output} \end{bmatrix} = \begin{bmatrix} .2 & .3 & .4 \\ .4 & .4 & .1 \\ .5 & .1 & .3 \end{bmatrix} \begin{bmatrix} \text{chemical input} \\ \text{food input} \\ \text{oil input} \end{bmatrix}.$$

Row 2 shows the inputs to produce food—a heavy use of chemicals and food, not so much oil. Row 3 of A shows the inputs consumed to refine a unit of oil. The real consumption matrix for the United States in 1958 contained 83 industries. The models in the 1990's are much larger and more precise. We chose a consumption matrix that has a convenient eigenvector.

Now comes the question: Can this economy meet demands y_1, y_2, y_3 for chemicals, food, and oil? To do that, the inputs p_1, p_2, p_3 will have to be higher—because part of $\boldsymbol{p}$ is consumed in producing $\boldsymbol{y}$. The input is $\boldsymbol{p}$ and the consumption is $A\boldsymbol{p}$, which leaves the output $\boldsymbol{p} - A\boldsymbol{p}$. This net production is what meets the demand $\boldsymbol{y}$:

Problem Find a vector $\boldsymbol{p}$ such that $\quad p - Ap = y \quad$ or $\quad p = (I - A)^{-1}y.$

Apparently the linear algebra question is whether $I - A$ is invertible. But there is more to the problem. The demand vector $\boldsymbol{y}$ is nonnegative, and so is A. *The production levels in* $\boldsymbol{p} = (I - A)^{-1}\boldsymbol{y}$ *must also be nonnegative.* The real question is:

When is $(I - A)^{-1}$ a nonnegative matrix?

This is the test on $(I - A)^{-1}$ for a productive economy, which can meet any positive demand. If A is small compared to I, then $A\boldsymbol{p}$ is small compared to $\boldsymbol{p}$. There is plenty of output. If A is too large, then production consumes more than it yields. In this case the external demand $\boldsymbol{y}$ cannot be met.

"Small" or "large" is decided by the largest eigenvalue λ_1 of A (which is positive):

If $\lambda_1 > 1$ then $(I - A)^{-1}$ has negative entries
If $\lambda_1 = 1$ then $(I - A)^{-1}$ fails to exist
If $\lambda_1 < 1$ then $(I - A)^{-1}$ is nonnegative as desired.

The main point is that last one. The reasoning uses a nice formula for $(I - A)^{-1}$, which we give now. The most important infinite series in mathematics is the ***geometric series*** $1 + x + x^2 + \cdots$. This series adds up to $1/(1 - x)$ provided x lies between -1 and 1. When $x = 1$ the series is $1 + 1 + 1 + \cdots = \infty$. When $|x| \geq 1$ the terms x^n don't go to zero and the series has no chance to converge.

The nice formula for $(I - A)^{-1}$ is the ***geometric series of matrices***:

Geometric series $\quad (I - A)^{-1} = I + A + A^2 + A^3 + \cdots.$

If you multiply the series $S = I + A + A^2 + \cdots$ by A, you get the same series except for I. Therefore $S - AS = I$, which is $(I - A)S = I$. The series adds to $S = (I - A)^{-1}$ if it converges. ***And it converges if all eigenvalues of A have*** $|\lambda| < 1$.

In our case $A \geq 0$. All terms of the series are nonnegative. Its sum is $(I - A)^{-1} \geq 0$.

Example 4 $A = \begin{bmatrix} .2 & .3 & .4 \\ .4 & .4 & .1 \\ .5 & .1 & .3 \end{bmatrix}$ has $\lambda_{\max} = .9$ and $(I - A)^{-1} = \frac{1}{93}\begin{bmatrix} 41 & 25 & 27 \\ 33 & 36 & 24 \\ 34 & 23 & 36 \end{bmatrix}$.

This economy is productive. A is small compared to I, because $\lambda_{\max}$ is .9. To meet the demand $\boldsymbol{y}$, start from $\boldsymbol{p} = (I - A)^{-1}\boldsymbol{y}$. Then $A\boldsymbol{p}$ is consumed in production, leaving $\boldsymbol{p} - A\boldsymbol{p}$. This is $(I - A)\boldsymbol{p} = \boldsymbol{y}$, and the demand is met.

Example 5 $A = \begin{bmatrix} 0 & 4 \\ 1 & 0 \end{bmatrix}$ has $\lambda_{\max} = 2$ and $(I - A)^{-1} = -\frac{1}{3}\begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix}$.

This consumption matrix A is too large. Demands can't be met, because production consumes more than it yields. The series $I + A + A^2 + \ldots$ does not converge to $(I - A)^{-1}$ because $\lambda_{\max} > 1$. The series is growing while $(I - A)^{-1}$ is actually negative.

In the same way $1 + 2 + 4 + \cdots$ is not really $1/(1 - 2) = -1$. But not entirely false !

Problem Set 8.3

Questions 1–12 are about Markov matrices and their eigenvalues and powers.

1 Find the eigenvalues of this Markov matrix (their sum is the trace):

$$A = \begin{bmatrix} .90 & .15 \\ .10 & .85 \end{bmatrix}.$$

What is the steady state eigenvector for the eigenvalue $\lambda_1 = 1$?

2 Diagonalize the Markov matrix in Problem 1 to $A = S\Lambda S^{-1}$ by finding its other eigenvector:

$$A = \begin{bmatrix} & \\ & \end{bmatrix}\begin{bmatrix} 1 & \\ & .75 \end{bmatrix}\begin{bmatrix} & \\ & \end{bmatrix}.$$

What is the limit of $A^k = S\Lambda^k S^{-1}$ when $\Lambda^k = \begin{bmatrix} 1 & 0 \\ 0 & .75^k \end{bmatrix}$ approaches $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$?

3 What are the eigenvalues and steady state eigenvectors for these Markov matrices?

$$A = \begin{bmatrix} 1 & .2 \\ 0 & .8 \end{bmatrix} \quad A = \begin{bmatrix} .2 & 1 \\ .8 & 0 \end{bmatrix} \quad A = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}.$$

4 For every 4 by 4 Markov matrix, what eigenvector of A^{T} corresponds to the (known) eigenvalue $\lambda = 1$?

5 Every year 2% of young people become old and 3% of old people become dead. (No births.) Find the steady state for

$$\begin{bmatrix}\text{young}\\ \text{old}\\ \text{dead}\end{bmatrix}_{k+1} = \begin{bmatrix}.98 & .00 & 0\\ .02 & .97 & 0\\ .00 & .03 & 1\end{bmatrix}\begin{bmatrix}\text{young}\\ \text{old}\\ \text{dead}\end{bmatrix}_k .$$

6 For a Markov matrix, the sum of the components of $\boldsymbol{x}$ equals the sum of the components of $A\boldsymbol{x}$. If $A\boldsymbol{x} = \lambda\boldsymbol{x}$ with $\lambda \neq 1$, prove that the components of this non-steady eigenvector $\boldsymbol{x}$ add to zero.

7 Find the eigenvalues and eigenvectors of A. Explain why A^k approaches A^∞:

$$A = \begin{bmatrix}.8 & .3\\ .2 & .7\end{bmatrix} \qquad A^\infty = \begin{bmatrix}.6 & .6\\ .4 & .4\end{bmatrix}.$$

Challenge problem: Which Markov matrices produce that steady state (.6, .4)?

8 The steady state eigenvector of a permutation matrix is $(\frac14, \frac14, \frac14, \frac14)$. This is *not* approached when $\boldsymbol{u}_0 = (0,0,0,1)$. What are $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ and $\boldsymbol{u}_3$ and $\boldsymbol{u}_4$? What are the four eigenvalues of P, which solve $\lambda^4 = 1$?

Permutation matrix = Markov matrix

$$P = \begin{bmatrix}0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\\ 1 & 0 & 0 & 0\end{bmatrix}.$$

9 Prove that the square of a Markov matrix is also a Markov matrix.

10 If $A = \left[\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right]$ is a Markov matrix, its eigenvalues are 1 and _____. The steady state eigenvector is $\boldsymbol{x}_1 =$ _____.

11 Complete A to a Markov matrix and find the steady state eigenvector. When A is a symmetric Markov matrix, why is $\boldsymbol{x}_1 = (1, \ldots, 1)$ its steady state?

$$A = \begin{bmatrix}.7 & .1 & .2\\ .1 & .6 & .3\\ - & - & -\end{bmatrix}.$$

12 A Markov differential equation is not $d\boldsymbol{u}/dt = A\boldsymbol{u}$ but $d\boldsymbol{u}/dt = (A - I)\boldsymbol{u}$. The diagonal is negative, the rest of $A - I$ is positive. The columns add to zero.

Find the eigenvalues of $B = A - I = \begin{bmatrix}-.2 & .3\\ .2 & -.3\end{bmatrix}$. Why does $A - I$ have $\lambda = 0$?

When $e^{\lambda_1 t}$ and $e^{\lambda_2 t}$ multiply $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, what is the steady state as $t \to \infty$?

Questions 13–15 are about linear algebra in economics.

13 Each row of the consumption matrix in Example 4 adds to .9. Why does that make $\lambda = .9$ an eigenvalue, and what is the eigenvector?

14 Multiply $I + A + A^2 + A^3 + \cdots$ by $I - A$ to show that the series adds to _____. For $A = \left[\begin{smallmatrix} 0 & \frac{1}{2} \\ 1 & 0 \end{smallmatrix}\right]$, find A^2 and A^3 and use the pattern to add up the series.

15 For which of these matrices does $I + A + A^2 + \cdots$ yield a nonnegative matrix $(I - A)^{-1}$? Then the economy can meet any demand:

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

If the demands are $y = (2, 6)$, what are the vectors $p = (I - A)^{-1} y$?

16 (Markov again) This matrix has zero determinant. What are its eigenvalues?

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

Find the limits of $A^k u_0$ starting from $u_0 = (1, 0, 0)$ and then $u_0 = (100, 0, 0)$.

17 If A is a Markov matrix, does $I + A + A^2 + \cdots$ add up to $(I - A)^{-1}$?

18 For the Leslie matrix show that $\det(A - \lambda I) = 0$ gives $F_1\lambda^2 + F_2P_1\lambda + F_3P_1P_2 = \lambda^3$. The right side λ^3 is larger as $\lambda \longrightarrow \infty$. The left side is larger at $\lambda = 1$ if $F_1 + F_2P_1 + F_3P_1P_2 > 1$. In that case the two sides are equal at an eigenvalue $\lambda_{\text{max}} > 1$: *growth.*

19 **Sensitivity of eigenvalues**: A matrix change ΔA produces eigenvalue changes $\Delta\Lambda$. *The formula for those changes* $\Delta\lambda_1, \ldots, \Delta\lambda_n$ *is* **diag**$(S^{-1}\,\Delta A\,S)$. **Challenge**: Start from $AS = S\Lambda$. The eigenvectors and eigenvalues change by ΔS and $\Delta\Lambda$:

$$(A + \Delta A)(S + \Delta S) = (S + \Delta S)(\Lambda + \Delta\Lambda) \text{ becomes } A(\Delta S) + (\Delta A)S = S(\Delta\Lambda) + (\Delta S)\Lambda.$$

Small terms $(\Delta A)(\Delta S)$ and $(\Delta S)(\Delta\Lambda)$ are ignored. *Multiply the last equation by* S^{-1}. From the inner terms, the diagonal part of $S^{-1}(\Delta A)S$ gives $\Delta\Lambda$ as we want. *Why do the outer terms* $S^{-1}\,A\,\Delta S$ *and* $S^{-1}\,\Delta S\,\Lambda$ *cancel on the diagonal?*

Explain $S^{-1}A = \Lambda S^{-1}$ and then **diag**$(\Lambda\,S^{-1}\,\Delta S) =$ **diag**$(S^{-1}\,\Delta S\,\Lambda)$.

20 Suppose $B > A > 0$, meaning that each $b_{ij} > a_{ij} > 0$. How does the Perron-Frobenius discussion show that $\lambda_{\text{max}}(B) > \lambda_{\text{max}}(A)$?

8.4 Linear Programming

Linear programming is linear algebra plus two new ideas: ***inequalities*** and ***minimization***. The starting point is still a matrix equation $Ax = b$. But the only acceptable solutions are *nonnegative*. We require $x \geq 0$ (meaning that no component of x can be negative). The matrix has $n > m$, more unknowns than equations. If there are any solutions $x \geq \mathbf{0}$ to $Ax = b$, there are probably a lot. Linear programming picks the solution $x^* \geq \mathbf{0}$ that minimizes the cost:

The cost is $c_1x_1 + \cdots + c_nx_n$. The winning vector x^* is the nonnegative solution of $Ax = b$ that has smallest cost.

Thus a linear programming problem starts with a matrix A and two vectors b and c:

i) A has $n > m$: for example $A = [1 \quad 1 \quad 2]$ (one equation, three unknowns)

ii) b has m components for m equations $Ax = b$: for example $b = [4]$

iii) The ***cost vector c*** has n components: for example $c = [5 \quad 3 \quad 8]$.

Then the problem is to minimize $c \cdot x$ subject to the requirements $Ax = b$ and $x \geq 0$:

$$\textbf{\textit{Minimize}} \quad 5x_1 + 3x_2 + 8x_3 \quad \textbf{\textit{subject to}} \quad x_1 + x_2 + 2x_3 = 4 \quad \textbf{\textit{and}} \quad x_1, x_2, x_3 \geq 0.$$

We jumped right into the problem, without explaining where it comes from. Linear programming is actually the most important application of mathematics to management. Development of the fastest algorithm and fastest code is highly competitive. You will see that finding x^* is harder than solving $Ax = b$, because of the extra requirements: $x^* \geq 0$ and minimum cost $c^{\mathrm{T}}x^*$. We will explain the background, and the famous *simplex method*, and *interior point methods*, after solving the example.

Look first at the "constraints": $Ax = b$ and $x \geq 0$. The equation $x_1 + x_2 + 2x_3 = 4$ gives a plane in three dimensions. The nonnegativity $x_1 \geq 0, x_2 \geq 0, x_3 \geq 0$ chops the plane down to a triangle. The solution x^* must lie in the triangle PQR in Figure 8.6.

Inside that triangle, all components of x are positive. On the edges of PQR, one component is zero. At the corners P and Q and R, two components are zero. ***The optimal solution x^* will be one of those corners!*** We will now show why.

The triangle contains all vectors x that satisfy $Ax = b$ and $x \geq 0$. Those x's are called *feasible points*, and the triangle is the ***feasible set***. These points are the allowed candidates in the minimization of $c \cdot x$, which is the final step:

Find x^* in the triangle PQR to minimize the cost $5x_1 + 3x_2 + 8x_3$.

The vectors that have *zero* cost lie on the plane $5x_1 + 3x_2 + 8x_3 = 0$. That plane does not meet the triangle. We cannot achieve zero cost, while meeting the requirements on x. So increase the cost C until the plane $5x_1 + 3x_2 + 8x_3 = C$ does meet the triangle. As C increases, we have *parallel planes moving toward the triangle*.

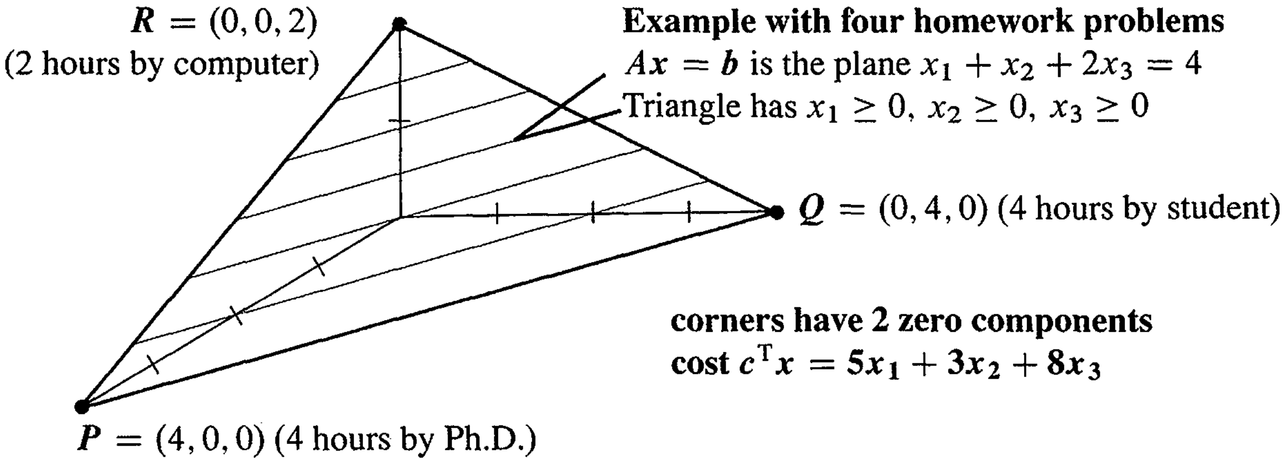

$R = (0,0,2)$
(2 hours by computer)
Example with four homework problems
$Ax = b$ is the plane $x_1 + x_2 + 2x_3 = 4$
Triangle has $x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0$
$Q = (0,4,0)$ (4 hours by student)
corners have 2 zero components
cost $c^{\mathrm{T}}x = 5x_1 + 3x_2 + 8x_3$
$P = (4,0,0)$ (4 hours by Ph.D.)

Figure 8.6: The triangle contains all nonnegative solutions: $Ax = b$ and $x \geq 0$. The lowest cost solution x^* is a corner P, Q, or R of this feasible set.

The first plane $5x_1 + 3x_2 + 8x_3 = C$ to touch the triangle has minimum cost C. ***The point where it touches is the solution x^*.*** This touching point must be one of the corners P or Q or R. A moving plane could not reach the inside of the triangle before it touches a corner! So check the cost $5x_1 + 3x_2 + 8x_3$ at each corner:

$$P = (4,0,0) \text{ costs } 20 \qquad Q = (0,4,0) \text{ costs } 12 \qquad R = (0,0,2) \text{ costs } 16.$$

The winner is Q. Then $x^* = (0,4,0)$ solves the linear programming problem.

If the cost vector c is changed, the parallel planes are tilted. For small changes, Q is still the winner. For the cost $c \cdot x = 5x_1 + 4x_2 + 7x_3$, the optimum x^* moves to $R = (0,0,2)$. The minimum cost is now $7 \cdot 2 = 14$.

Note 1 Some linear programs *maximize profit* instead of minimizing cost. The mathematics is almost the same. The parallel planes start with a large value of C, instead of a small value. They move toward the origin (instead of away), as C gets smaller. *The first touching point is still a corner.*

Note 2 The requirements $Ax = b$ and $x \geq 0$ could be impossible to satisfy. The equation $x_1 + x_2 + x_3 = -1$ cannot be solved with $x \geq 0$. *That feasible set is empty.*

Note 3 It could also happen that the feasible set is *unbounded.* If the requirement is $x_1 + x_2 - 2x_3 = 4$, the large positive vector $(100, 100, 98)$ is now a candidate. So is the larger vector $(1000, 1000, 998)$. The plane $Ax = b$ is no longer chopped off to a triangle. The two corners P and Q are still candidates for x^*, but R moved to infinity.

Note 4 With an unbounded feasible set, the minimum cost could be $-\infty$ (*minus infinity*). Suppose the cost is $-x_1 - x_2 + x_3$. Then the vector $(100, 100, 98)$ costs $C = -102$. The vector $(1000, 1000, 998)$ costs $C = -1002$. We are being paid to include x_1 and x_2, instead of paying a cost. In realistic applications this will not happen. But it is theoretically possible that A, b, and c can produce unexpected triangles and costs.

The Primal and Dual Problems

This first problem will fit $\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}$ in that example. The unknowns x_1, x_2, x_3 represent hours of work by a Ph.D. and a student and a machine. The costs per hour are \$5, \$3, and \$8. (*I apologize for such low pay.*) The number of hours cannot be negative: $x_1 \geq 0, x_2 \geq 0, x_3 \geq 0$. The Ph.D. and the student get through one homework problem per hour. *The machine solves two problems in one hour.* In principle they can share out the homework, which has four problems to be solved: $x_1 + x_2 + 2x_3 = 4$.

The problem is to finish the four problems at minimum cost $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$.

If all three are working, the job takes one hour: $x_1 = x_2 = x_3 = 1$. The cost is $5 + 3 + 8 = 16$. But certainly the Ph.D. should be put out of work by the student (who is just as fast and costs less—this problem is getting realistic). When the student works two hours and the machine works one, the cost is $6 + 8$ and all four problems get solved. We are on the edge $\boldsymbol{QR}$ of the triangle because the Ph.D. is not working: $x_1 = 0$. But the best point is all work by student (at $\boldsymbol{Q}$) or all work by machine (at $\boldsymbol{R}$). In this example the student solves four problems in four hours for \$12—the minimum cost.

With only one equation in $Ax = \boldsymbol{b}$, the corner $(0, 4, 0)$ has only one nonzero component. ***When $Ax = \boldsymbol{b}$ has m equations, corners have m nonzeros.*** We solve $Ax = \boldsymbol{b}$ for those m variables, with $n - m$ free variables set to zero. But unlike Chapter 3, ***we don't know which m variables to choose.***

The number of possible corners is the number of ways to choose m components out of n. This number "n choose m" is heavily involved in gambling and probability. With $n = 20$ unknowns and $m = 8$ equations (still small numbers), the "feasible set" can have $20!/8!12!$ corners. That number is $(20)(19)\cdots(13) = 5{,}079{,}110{,}400$.

Checking three corners for the minimum cost was fine. Checking five billion corners is not the way to go. The simplex method described below is much faster.

The Dual Problem In linear programming, problems come in pairs. There is a minimum problem and a maximum problem—the original and its "dual." The original problem was specified by a matrix A and two vectors $\boldsymbol{b}$ and $\boldsymbol{c}$. The dual problem transposes A and switches $\boldsymbol{b}$ and $\boldsymbol{c}$: ***Maximize $\boldsymbol{b} \cdot \boldsymbol{y}$***. Here is the dual to our example:

> **A cheater offers to solve homework problems by selling the answers.** The charge is y dollars per problem, or $4y$ altogether. (Note how $\boldsymbol{b} = 4$ has gone into the cost.) The cheater must be as cheap as the Ph.D. or student or machine: $y \leq 5$ and $y \leq 3$ and $2y \leq 8$. (Note how $\boldsymbol{c} = (5, 3, 8)$ has gone into inequality constraints). The cheater maximizes the income $4y$.

Dual Problem ***Maximize $\boldsymbol{b} \cdot \boldsymbol{y}$ subject to $A^{\mathrm{T}}\boldsymbol{y} \leq \boldsymbol{c}$.***

The maximum occurs when $y = 3$. The income is $4y = 12$. The maximum in the dual problem (\$12) equals the minimum in the original (\$12). *Max = min* is duality.

If either problem has a best vector (x^* or y^*) then so does the other.
Minimum cost $c \cdot x^*$ equals maximum income $b \cdot y^*$

This book started with a row picture and a column picture. The first "duality theorem" was about rank: The number of independent rows equals the number of independent columns. That theorem, like this one, was easy for small matrices. Minimum cost = maximum income is proved in our text *Linear Algebra and Its Applications*. One line will establish the easy half of the theorem: ***The cheater's income $b^{\mathrm{T}}y$ cannot exceed the honest cost:***

$$\textit{If } Ax = b, x \geq 0, A^{\mathrm{T}}y \leq c \quad \textit{then} \quad b^{\mathrm{T}}y = (Ax)^{\mathrm{T}}y = x^{\mathrm{T}}(A^{\mathrm{T}}y) \leq x^{\mathrm{T}}c. \qquad (1)$$

The full duality theorem says that when $b^{\mathrm{T}}y$ reaches its maximum and $x^{\mathrm{T}}c$ reaches its minimum, they are equal: $b \cdot y^* = c \cdot x^*$. Look at the last step in (1), with $\leq$ sign:

The dot product of $x \geq 0$ and $s = c - A^{\mathrm{T}}y \geq 0$ gave $x^{\mathrm{T}}s \geq 0$. This is $x^{\mathrm{T}}A^{\mathrm{T}}y \leq x^{\mathrm{T}}c$.

Equality needs $x^{\mathrm{T}}s = 0$ So the optimal solution has $x_j^* = 0$ or $s_j^* = 0$ for each j.

The Simplex Method

Elimination is the workhorse for linear equations. The simplex method is the workhorse for linear inequalities. We cannot give the simplex method as much space as elimination, but the idea can be clear. *The simplex method goes from one corner to a neighboring corner of lower cost.* Eventually (and quite soon in practice) it reaches the corner of minimum cost.

A ***corner*** is a vector $x \geq 0$ that satisfies the m equations $Ax = b$ with at most m positive components. *The other $n - m$ components are zero.* (Those are the free variables. Back substitution gives the m basic variables. All variables must be nonnegative or x is a false corner.) For a *neighboring corner*, one zero component of x becomes positive and one positive component becomes zero.

The simplex method must decide which component "enters" by becoming positive, and which component "leaves" by becoming zero. That exchange is chosen so as to lower the total cost. This is one step of the simplex method, moving toward x^*.

Here is the overall plan. Look at each zero component at the current corner. If it changes from 0 to 1, the other nonzeros have to adjust to keep $Ax = b$. Find the new x by back substitution and compute the change in the total cost $c \cdot x$. This change is the "reduced cost" r of the new component. The ***entering variable*** is the one that gives the ***most negative*** r. This is the greatest cost reduction for a single unit of a new variable.

Example 1 Suppose the current corner is $P = (4, 0, 0)$, with the Ph.D. doing all the work (the cost is \$20). If the student works one hour, the cost of $x = (3, 1, 0)$ is down to \$18. The reduced cost is $r = -2$. If the machine works one hour, then $x = (2, 0, 1)$ also costs \$18. The reduced cost is also $r = -2$. In this case the simplex method can choose either the student or the machine as the entering variable.

Even in this small example, the first step may not go immediately to the best $\boldsymbol{x}^*$. The method chooses the entering variable before it knows how much of that variable to include. We computed r when the entering variable changes from 0 to 1, but one unit may be too much or too little. The method now chooses the leaving variable (the Ph.D.). It moves to corner $\boldsymbol{Q}$ or $\boldsymbol{R}$ in the figure.

The more of the entering variable we include, the lower the cost. This has to stop when one of the positive components (which are adjusting to keep $A\boldsymbol{x} = \boldsymbol{b}$) hits zero. ***The leaving variable** is the first positive x_i to reach zero.* When that happens, a neighboring corner has been found. Then start again (from the new corner) to find the next variables to enter and leave.

When all reduced costs are positive, the current corner is the optimal $\boldsymbol{x}^*$. No zero component can become positive without increasing $\boldsymbol{c} \cdot \boldsymbol{x}$. No new variable should enter. The problem is solved (and we can show that $\boldsymbol{y}^*$ is found too).

Note Generally $\boldsymbol{x}^*$ is reached in αn steps, where α is not large. But examples have been invented which use an exponential number of simplex steps. Eventually a different approach was developed, which is guaranteed to reach $\boldsymbol{x}^*$ in fewer (but more difficult) steps. The new methods travel through the *interior* of the feasible set.

Example 2 Minimize the cost $\boldsymbol{c} \cdot \boldsymbol{x} = 3x_1 + x_2 + 9x_3 + x_4$. The constraints are $\boldsymbol{x} \geq 0$ and two equations $A\boldsymbol{x} = \boldsymbol{b}$:

$$\begin{aligned} x_1 + 2x_3 + x_4 &= 4 \qquad & m &= 2 \quad \text{equations} \\ x_2 + x_3 - x_4 &= 2 \qquad & n &= 4 \quad \text{unknowns.} \end{aligned}$$

A starting corner is $\boldsymbol{x} = (4, 2, 0, 0)$ which costs $\boldsymbol{c} \cdot \boldsymbol{x} = 14$. It has $m = 2$ nonzeros and $n - m = 2$ zeros. The zeros are x_3 and x_4. The question is whether x_3 or x_4 should enter (become nonzero). Try one unit of each of them:

$$\begin{aligned} &\text{If } x_3 = 1 \text{ and } x_4 = 0, \quad \text{then } \boldsymbol{x} = (2, 1, 1, 0) \text{ costs } 16. \\ &\text{If } x_4 = 1 \text{ and } x_3 = 0, \quad \text{then } \boldsymbol{x} = (3, 3, 0, 1) \text{ costs } 13. \end{aligned}$$

Compare those costs with 14. The reduced cost of x_3 is $r = 2$, positive and useless. The reduced cost of x_4 is $r = -1$, negative and helpful. *The entering variable is x_4.*

How much of x_4 can enter? One unit of x_4 made x_1 drop from 4 to 3. Four units will make x_1 drop from 4 to zero (while x_2 increases all the way to 6). *The leaving variable is x_1.* The new corner is $\boldsymbol{x} = (0, 6, 0, 4)$, which costs only $\boldsymbol{c} \cdot \boldsymbol{x} = 10$. This is the optimal $\boldsymbol{x}^*$, but to know that we have to try another simplex step from $(0, 6, 0, 4)$. Suppose x_1 or x_3 tries to enter:

$$\begin{array}{lll} \textbf{Start from the} & \text{If } x_1 = 1 \text{ and } x_3 = 0, & \text{then } \boldsymbol{x} = (1, 5, 0, 3) \text{ costs } 11. \\ \textbf{corner } (\mathbf{0}, \mathbf{6}, \mathbf{0}, \mathbf{4}) & \text{If } x_3 = 1 \text{ and } x_1 = 0, & \text{then } \boldsymbol{x} = (0, 3, 1, 2) \text{ costs } 14. \end{array}$$

Those costs are higher than 10. Both r's are positive—it does not pay to move. The current corner $(0, 6, 0, 4)$ is the solution $\boldsymbol{x}^*$.

These calculations can be streamlined. Each simplex step solves three linear systems with the same matrix B. (This is the m by m matrix that keeps the m basic columns of A.) When a column enters and an old column leaves, there is a quick way to update B^{-1}. That is how most codes organize the simplex method.

Our text on *Computational Science and Engineering* includes a short code with comments. (The code is also on **math.mit.edu/cse**) The best y^* solves m equations $A^{\mathrm{T}}y^* = c$ in the m components that are nonzero in x^*. Then we have optimality $x^{\mathrm{T}}s = 0$ and this is duality: *Either* $x_j^* = 0$ *or the "slack" in* $s^* = c - A^{\mathrm{T}}y^*$ *has* $s_j^* = 0$.

When $x^* = (0, 4, 0)$ was the optimal corner Q, the cheater's price was set by $y^* = 3$.

Interior Point Methods

The simplex method moves along the edges of the feasible set, eventually reaching the optimal corner x^*. ***Interior point methods move inside the feasible set*** (where $x > \mathbf{0}$). These methods hope to go more directly to x^*. They work well.

One way to stay inside is to put a barrier at the boundary. Add extra cost as a *logarithm that blows up* when any variable x_j touches zero. The best vector has $x > \mathbf{0}$. The number θ is a small parameter that we move toward zero.

Barrier problem **Minimize** $c^{\mathrm{T}}x - \theta\,(\log x_1 + \cdots + \log x_n)$ **with** $Ax = b$ (2)

This cost is nonlinear (but linear programming is already nonlinear from inequalities). The constraints $x_j \geq 0$ are not needed because $\log x_j$ becomes infinite at $x_j = 0$.

The barrier gives an *approximate problem* for each θ. The m constraints $Ax = b$ have Lagrange multipliers $y_1, \ldots, y_m$. This is the good way to deal with constraints.

y from Lagrange $$L(x, y, \theta) = c^{\mathrm{T}}x - \theta\,(\textstyle\sum \log x_i) - y^{\mathrm{T}}(Ax - b) \qquad (3)$$

$\partial L/\partial y = 0$ brings back $Ax = b$. The derivatives $\partial L/\partial x_j$ are interesting!

Optimality in barrier pbm $$\frac{\partial L}{\partial x_j} = c_j - \frac{\theta}{x_j} - (A^{\mathrm{T}}y)_j = 0 \quad \text{which is} \quad x_j s_j = \theta\,. \qquad (4)$$

The true problem has $x_j s_j = 0$. The barrier problem has $x_j s_j = \theta$. The solutions $x^*(\theta)$ lie on the ***central path*** to $x^*(0)$. Those n optimality equations $x_j s_j = \theta$ are nonlinear, and we solve them iteratively by Newton's method.

The current x, y, s will satisfy $Ax = b, x \geq \mathbf{0}$ and $A^{\mathrm{T}}y + s = c$, *but not* $x_j s_j = \theta$. Newton's method takes a step $\Delta x, \Delta y, \Delta s$. By ignoring the second-order term $\Delta x \Delta s$ in $(x + \Delta x)(s + \Delta s) = \theta$, the corrections in x, y, s come from linear equations:

Newton step $$\begin{aligned} A\,\Delta x &= 0 \\ A^{\mathrm{T}}\Delta y + \Delta s &= 0 \\ s_j \Delta x_j + x_j \Delta s_j &= \theta - x_j s_j \end{aligned} \qquad (5)$$

Newton iteration has quadratic convergence for each θ, and then θ approaches zero. The duality gap $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{s}$ generally goes below 10^{-8} after 20 to 60 steps. The explanation in my *Computational Science and Engineering* textbook takes one Newton step in detail, for the example with four homework problems. I didn't intend that the student should end up doing all the work, but $\boldsymbol{x}^*$ turned out that way.

This interior point method is used almost "as is" in commercial software, for a large class of linear and nonlinear optimization problems.

Problem Set 8.4

1 Draw the region in the xy plane where $x + 2y = 6$ and $x \geq 0$ and $y \geq 0$. Which point in this "feasible set" minimizes the cost $c = x + 3y$? Which point gives maximum cost? Those points are at corners.

2 Draw the region in the xy plane where $x + 2y \leq 6$, $2x + y \leq 6$, $x \geq 0$, $y \geq 0$. It has four corners. Which corner minimizes the cost $c = 2x - y$?

3 What are the corners of the set $x_1 + 2x_2 - x_3 = 4$ with x_1, x_2, x_3 all ≥ 0? Show that the cost $x_1 + 2x_3$ can be very negative in this feasible set. This is an example of unbounded cost: no minimum.

4 Start at $\boldsymbol{x} = (0, 0, 2)$ where the machine solves all four problems for \$16. Move to $\boldsymbol{x} = (0, 1, \quad)$ to find the reduced cost r (the savings per hour) for work by the student. Find r for the Ph.D. by moving to $\boldsymbol{x} = (1, 0, \quad)$ with 1 hour of Ph.D. work.

5 Start Example 1 from the Ph.D. corner $(4, 0, 0)$ with $\boldsymbol{c}$ changed to $[\,5 \quad 3 \quad 7\,]$. Show that r is better for the machine even when the total cost is lower for the student. The simplex method takes two steps, first to the machine and then to the student for $\boldsymbol{x}^*$.

6 Choose a different cost vector $\boldsymbol{c}$ so the Ph.D. gets the job. Rewrite the dual problem (maximum income to the cheater).

7 A six-problem homework on which the Ph.D. is fastest gives a second constraint $2x_1 + x_2 + x_3 = 6$. Then $\boldsymbol{x} = (2, 2, 0)$ shows two hours of work by Ph.D. and student on each homework. Does this $\boldsymbol{x}$ minimize the cost $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$ with $\boldsymbol{c} = (5, 3, 8)$?

8 These two problems are also dual. Prove weak duality, that always $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b} \leq \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$:

Primal problem Minimize $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$ with $A\boldsymbol{x} \geq \boldsymbol{b}$ and $\boldsymbol{x} \geq \boldsymbol{0}$.
Dual problem Maximize $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b}$ with $A^{\mathrm{T}}\boldsymbol{y} \leq \boldsymbol{c}$ and $\boldsymbol{y} \geq \boldsymbol{0}$.

8.5 Fourier Series: Linear Algebra for Functions

This section goes from finite dimensions to *infinite* dimensions. I want to explain linear algebra in infinite-dimensional space, and to show that it still works. First step: look back. This book began with vectors and dot products and linear combinations. We begin by converting those basic ideas to the infinite case—then the rest will follow.

What does it mean for a vector to have infinitely many components? There are two different answers, both good:

1. The vector becomes $\boldsymbol{v} = (v_1, v_2, v_3, \ldots)$. It could be $(1, \frac{1}{2}, \frac{1}{4}, \ldots)$.

2. The vector becomes a function $f(x)$. It could be $\sin x$.

We will go both ways. Then the idea of Fourier series will connect them.

After vectors come *dot products.* The natural dot product of two infinite vectors $(v_1, v_2, \ldots)$ and $(w_1, w_2, \ldots)$ is an infinite series:

$$\textbf{Dot product} \qquad \boldsymbol{v} \cdot \boldsymbol{w} = v_1 w_1 + v_2 w_2 + \cdots. \tag{1}$$

This brings a new question, which never occurred to us for vectors in $\mathbf{R}^n$. Does this infinite sum add up to a finite number? Does the series converge? Here is the first and biggest difference between finite and infinite.

When $\boldsymbol{v} = \boldsymbol{w} = (1, 1, 1, \ldots)$, the sum certainly does not converge. In that case $\boldsymbol{v} \cdot \boldsymbol{w} = 1 + 1 + 1 + \cdots$ is infinite. Since $\boldsymbol{v}$ equals $\boldsymbol{w}$, we are really computing $\boldsymbol{v} \cdot \boldsymbol{v} = \|\boldsymbol{v}\|^2 =$ length squared. The vector $(1, 1, 1, \ldots)$ has infinite length. *We don't want that vector.* Since we are making the rules, we don't have to include it. The only vectors to be allowed are those with finite length:

DEFINITION The vector $(v_1, v_2, \ldots)$ is in our infinite-dimensional ***"Hilbert space"*** if and only if its length $\|\boldsymbol{v}\|$ is finite:

$$\|\boldsymbol{v}\|^2 = \boldsymbol{v} \cdot \boldsymbol{v} = v_1^2 + v_2^2 + v_3^2 + \cdots \text{ must add to a finite number.}$$

Example 1 The vector $\boldsymbol{v} = (1, \frac{1}{2}, \frac{1}{4}, \ldots)$ is included in Hilbert space, because its length is $2/\sqrt{3}$. We have a geometric series that adds to $4/3$. The length of $\boldsymbol{v}$ is the square root:

$$\textbf{Length squared} \qquad \boldsymbol{v} \cdot \boldsymbol{v} = 1 + \tfrac{1}{4} + \tfrac{1}{16} + \cdots = \frac{1}{1 - \frac{1}{4}} = \tfrac{4}{3}.$$

Question If $\boldsymbol{v}$ and $\boldsymbol{w}$ have finite length, how large can their dot product be?

Answer The sum $\boldsymbol{v} \cdot \boldsymbol{w} = v_1 w_1 + v_2 w_2 + \cdots$ also adds to a finite number. We can safely take dot products. The Schwarz inequality is still true:

$$\textbf{Schwarz inequality} \qquad |\boldsymbol{v} \cdot \boldsymbol{w}| \le \|\boldsymbol{v}\| \, \|\boldsymbol{w}\|. \tag{2}$$

The ratio of $\boldsymbol{v} \cdot \boldsymbol{w}$ to $\|\boldsymbol{v}\| \, \|\boldsymbol{w}\|$ is still the cosine of θ (the angle between $\boldsymbol{v}$ and $\boldsymbol{w}$). Even in infinite-dimensional space, $|\cos \theta|$ is not greater than 1.

Now change over to functions. Those are the "vectors." The space of functions $f(x)$, $g(x), h(x), \ldots$ defined for $0 \leq x \leq 2\pi$ must be somehow bigger than $\mathbf{R}^n$. ***What is the dot product of $f(x)$ and $g(x)$? What is the length of $f(x)$?***

Key point in the continuous case: *Sums are replaced by integrals.* Instead of a sum of v_j times w_j, the dot product is an integral of $f(x)$ times $g(x)$. Change the "dot" to parentheses with a comma, and change the words "dot product" to *inner product*:

DEFINITION The ***inner product*** of $f(x)$ and $g(x)$, and the ***length squared***, are

$$(f, g) = \int_0^{2\pi} f(x)g(x)\,dx \quad \text{and} \quad \|f\|^2 = \int_0^{2\pi} (f(x))^2\,dx. \tag{3}$$

The interval $[0, 2\pi]$ where the functions are defined could change to a different interval like $[0, 1]$ or $(-\infty, \infty)$. We chose 2π because our first examples are $\sin x$ and $\cos x$.

Example 2 The length of $f(x) = \sin x$ comes from its inner product with itself:

$$(f, f) = \int_0^{2\pi} (\sin\ x)^2\,dx = \pi. \quad \text{The length of } \sin x \text{ is } \sqrt{\pi}.$$

That is a standard integral in calculus—not part of linear algebra. By writing $\sin^2 x$ as $\frac{1}{2} - \frac{1}{2}\cos 2x$, we see it go above and below its average value $\frac{1}{2}$. Multiply that average by the interval length 2π to get the answer π.

More important: $\sin x$ ***and*** $\cos x$ ***are orthogonal in function space***:

$$\begin{matrix}\textbf{Inner product}\\ \textbf{is zero}\end{matrix} \qquad \int_0^{2\pi} \sin x \cos x\,dx = \int_0^{2\pi} \tfrac{1}{2}\sin 2x\,dx = \left[-\tfrac{1}{4}\cos 2x\right]_0^{2\pi} = 0. \tag{4}$$

This zero is no accident. It is highly important to science. The orthogonality goes beyond the two functions $\sin x$ and $\cos x$, to an infinite list of sines and cosines. The list contains $\cos 0x$ (which is 1), $\sin x$, $\cos x, \sin 2x, \cos 2x, \sin 3x, \cos 3x, \ldots$..

Every function in that list is orthogonal to every other function in the list.

Fourier Series

The Fourier series of a function $y(x)$ is its expansion into sines and cosines:

$$y(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \cdots. \tag{5}$$

We have an orthogonal basis! The vectors in "function space" are combinations of the sines and cosines. On the interval from $x = 2\pi$ to $x = 4\pi$, all our functions repeat what they did from 0 to 2π. They are "*periodic*." The distance between repetitions is the period 2π.

Remember: The list is infinite. The Fourier series is an infinite series. We avoided the vector $v = (1, 1, 1, \ldots)$ because its length is infinite, now we avoid a function like $\frac{1}{2} + \cos x + \cos 2x + \cos 3x + \cdots$. (*Note*: This is π times the famous **delta function** $\delta(x)$. It is an infinite "spike" above a single point. At $x = 0$ its height $\frac{1}{2} + 1 + 1 + \cdots$ is infinite. At all points inside $0 < x < 2\pi$ the series adds in some average way to zero.) The integral of $\delta(x)$ is 1. But $\int \delta^2(x) = \infty$, so delta functions are excluded from Hilbert space.

Compute the length of a typical sum $f(x)$:

$$\begin{aligned}(f, f) &= \int_0^{2\pi} (a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + \cdots)^2 \, dx \\ &= \int_0^{2\pi} (a_0^2 + a_1^2 \cos^2 x + b_1^2 \sin^2 x + a_2^2 \cos^2 2x + \cdots) \, dx \\ \|f\|^2 &= 2\pi a_0^2 + \pi(a_1^2 + b_1^2 + a_2^2 + \cdots). \end{aligned} \tag{6}$$

The step from line 1 to line 2 used orthogonality. All products like $\cos x \cos 2x$ integrate to give zero. Line 2 contains what is left—the integrals of each sine and cosine squared. Line 3 evaluates those integrals. (The integral of 1^2 is 2π, when all other integrals give π.) If we divide by their lengths, our functions become *orthonormal*:

$$\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \ldots \textit{ is an orthonormal basis for our function space.}$$

These are unit vectors. We could combine them with coefficients $A_0, A_1, B_1, A_2, \ldots$ to yield a function $F(x)$. Then the 2π and the π's drop out of the formula for length.

Function length = vector length $$\|F\|^2 = (F, F) = A_0^2 + A_1^2 + B_1^2 + A_2^2 + \cdots . \tag{7}$$

Here is the important point, for $f(x)$ as well as $F(x)$. *The function has finite length exactly when the vector of coefficients has finite length.* Fourier series gives us a perfect match between function space and infinite-dimensional Hilbert space. The function is in L^2, its Fourier coefficients are in ℓ^2.

The function space contains $f(x)$ exactly when the Hilbert space contains the vector $v = (a_0, a_1, b_1, \ldots)$ of Fourier coefficients. Both $f(x)$ and v have finite length.

Example 3 Suppose $f(x)$ is a "square wave," equal to 1 for $0 \le x < \pi$. Then $f(x)$ drops to -1 for $\pi \le x < 2\pi$. The $+1$ and -1 repeats forever. This $f(x)$ is an odd function like the sines, and all its cosine coefficients are zero. We will find its Fourier series, containing only sines:

Square wave $$f(x) = \frac{4}{\pi}\left[\frac{\sin x}{1} + \frac{\sin 3x}{3} + \frac{\sin 5x}{5} + \cdots\right]. \tag{8}$$

The length is $\sqrt{2\pi}$, because at every point $(f(x))^2$ is $(-1)^2$ or $(+1)^2$:

$$\|f\|^2 = \int_0^{2\pi} (f(x))^2 \, dx = \int_0^{2\pi} 1 \, dx = 2\pi.$$

At $x = 0$ the sines are zero and the Fourier series gives zero. This is half way up the jump from -1 to $+1$. The Fourier series is also interesting when $x = \frac{\pi}{2}$. At this point the square wave equals 1, and the sines in (8) alternate between $+1$ and -1:

$$\textbf{Formula for } \pi \qquad 1 = \frac{4}{\pi}\left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots\right). \tag{9}$$

Multiply by π to find a magical formula $4(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots)$ for that famous number.

The Fourier Coefficients

How do we find the a's and b's which multiply the cosines and sines? For a given function $f(x)$, we are asking for its Fourier coefficients:

$$\textbf{Fourier series} \qquad f(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + \cdots .$$

Here is the way to find a_1. ***Multiply both sides by* $\cos x$*. Then integrate from* 0 *to* 2π.** The key is orthogonality! All integrals on the right side are zero, except for $\cos^2 x$:

$$\textbf{Coefficient } a_1 \qquad \int_0^{2\pi} f(x) \cos x \, dx = \int_0^{2\pi} a_1 \cos^2 x \, dx = \pi a_1. \tag{10}$$

Divide by π and you have a_1. To find any other a_k, multiply the Fourier series by $\cos kx$. Integrate from 0 to 2π. Use orthogonality, so only the integral of $a_k \cos^2 kx$ is left. That integral is πa_k, and divide by π:

$$a_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos kx \, dx \quad \text{and similarly} \quad b_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin kx \, dx. \tag{11}$$

The exception is a_0. This time we multiply by $\cos 0x = 1$. The integral of 1 is 2π:

$$\textbf{Constant term} \quad a_0 = \frac{1}{2\pi} \int_0^{2\pi} f(x) \cdot 1 \, dx = \textbf{\textit{average value}} \text{ of } f(x). \tag{12}$$

I used those formulas to find the Fourier coefficients for the square wave. The integral of $f(x) \cos kx$ was zero. The integral of $f(x) \sin kx$ was $4/k$ for odd k.

Compare Linear Algebra in $\mathbf{R}^n$

The point to emphasize is how this infinite-dimensional case is so much like the n-dimensional case. Suppose the nonzero vectors $v_1, \ldots, v_n$ are orthogonal. We want to write the vector b (instead of the function $f(x)$) as a combination of those v's:

$$\textbf{Finite orthogonal series} \quad b = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n. \tag{13}$$

Multiply both sides by v_1^{T}. Use orthogonality, so $v_1^{\mathrm{T}} v_2 = 0$. Only the c_1 term is left:

$$\textbf{Coefficient } c_1 \quad v_1^{\mathrm{T}} b = c_1 v_1^{\mathrm{T}} v_1 + 0 + \cdots + 0. \quad \text{Therefore } c_1 = v_1^{\mathrm{T}} b / v_1^{\mathrm{T}} v_1. \tag{14}$$

The denominator $v_1^{\mathrm{T}} v_1$ is the length squared, like π in equation 11. The numerator $v_1^{\mathrm{T}} b$ is the inner product like $\int f(x) \cos kx \, dx$. ***Coefficients are easy to find when the basis***

vectors are orthogonal. We are just doing one-dimensional projections, to find the components along each basis vector.

The formulas are even better when the vectors are orthonormal. Then we have unit vectors. The denominators $v_k^{\mathrm{T}} v_k$ are all 1. You know $c_k = v_k^{\mathrm{T}} \boldsymbol{b}$ in another form:

$$\textbf{Equation for } \boldsymbol{c}\textbf{'s} \quad c_1 v_1 + \cdots + c_n v_n = \boldsymbol{b} \quad \text{or} \quad \begin{bmatrix} & & \\ v_1 & \cdots & v_n \\ & & \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \boldsymbol{b}.$$

The $\boldsymbol{v}$'s are in an orthogonal matrix Q. Its inverse is Q^{T}. That gives the c's:

$$Q\boldsymbol{c} = \boldsymbol{b} \quad \text{yields} \quad \boldsymbol{c} = Q^{\mathrm{T}}\boldsymbol{b}. \quad \text{Row by row this is } c_k = \boldsymbol{q}_k^{\mathrm{T}}\boldsymbol{b}.$$

Fourier series is like having a matrix with infinitely many orthogonal columns. Those columns are the basis functions $1, \cos x, \sin x, \ldots$. After dividing by their lengths we have an "infinite orthogonal matrix." Its inverse is its transpose. Orthogonality is what reduces a series of terms to one single term.

Problem Set 8.5

1 Integrate the trig identity $2\cos jx \cos kx = \cos(j+k)x + \cos(j-k)x$ to show that $\cos jx$ is orthogonal to $\cos kx$, provided $j \neq k$. What is the result when $j = k$?

2 Show that $1, x$, and $x^2 - \frac{1}{3}$ are orthogonal, when the integration is from $x = -1$ to $x = 1$. Write $f(x) = 2x^2$ as a combination of those orthogonal functions.

3 Find a vector $(w_1, w_2, w_3, \ldots)$ that is orthogonal to $\boldsymbol{v} = (1, \frac{1}{2}, \frac{1}{4}, \ldots)$. Compute its length $\|\boldsymbol{w}\|$.

4 The first three *Legendre polynomials* are $1, x$, and $x^2 - \frac{1}{3}$. Choose c so that the fourth polynomial $x^3 - cx$ is orthogonal to the first three. All integrals go from -1 to 1.

5 For the square wave $f(x)$ in Example 3, show that

$$\int_0^{2\pi} f(x)\cos x\,dx = 0 \qquad \int_0^{2\pi} f(x)\sin x\,dx = 4 \qquad \int_0^{2\pi} f(x)\sin 2x\,dx = 0.$$

Which three Fourier coefficients come from those integrals?

6 The square wave has $\|f\|^2 = 2\pi$. Then (6) gives what remarkable sum for π^2?

7 Graph the square wave. Then graph by hand the sum of two sine terms in its series, or graph by machine the sum of 2, 3, and 10 terms. The famous ***Gibbs phenomenon*** is the oscillation that overshoots the jump (this doesn't die down with more terms).

8 Find the lengths of these vectors in Hilbert space:

(a) $\boldsymbol{v} = \left(\frac{1}{\sqrt{1}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{4}}, \ldots\right)$

(b) $v = (1, a, a^2, \ldots)$

(c) $f(x) = 1 + \sin x$.

9 Compute the Fourier coefficients a_k and b_k for $f(x)$ defined from 0 to 2π:

(a) $f(x) = 1$ for $0 \leq x \leq \pi$, $f(x) = 0$ for $\pi < x < 2\pi$

(b) $f(x) = x$.

10 When $f(x)$ has period 2π, why is its integral from $-\pi$ to π the same as from 0 to 2π? If $f(x)$ is an *odd* function, $f(-x) = -f(x)$, show that $\int_0^{2\pi} f(x)\,dx$ is zero. Odd functions only have sine terms, even functions have cosines.

11 From trig identities find the only two terms in the Fourier series for $f(x)$:

(a) $f(x) = \cos^2 x$ (b) $f(x) = \cos\left(x + \frac{\pi}{3}\right)$ (c) $f(x) = \sin^3 x$

12 The functions $1, \cos x, \sin x, \cos 2x, \sin 2x, \ldots$ are a basis for Hilbert space. Write the derivatives of those first five functions as combinations of the same five functions. What is the 5 by 5 "differentiation matrix" for these functions?

13 Find the Fourier coefficients a_k and b_k of the square pulse $F(x)$ centered at $x = 0$: $F(x) = 1/h$ for $|x| \leq h/2$ and $F(x) = 0$ for $h/2 < |x| \leq \pi$.

As $h \to 0$, this $F(x)$ approaches a delta function. Find the limits of a_k and b_k.

The Fourier Series section 4.1 of *Computational Science and Engineering* explains the sine series, cosine series, complete series, and complex series $\Sigma c_k e^{ikx}$ on **math.mit.edu/cse**.

8.6 Linear Algebra for Statistics and Probability

Statistics deals with data, often in large quantities. Since data tends to go into rectangular matrices, we expect to see $A^{\mathrm{T}}A$. The least squares problem $A\widehat{x} \approx \boldsymbol{b}$ is ***linear regression.*** The best solution $\widehat{x}$ fits m observations by $n < m$ parameters. This is a fundamental application of linear algebra to statistics.

This section goes beyond $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}\boldsymbol{b}$. These unweighted equations assume that the measurements $b_1, \ldots, b_m$ are equally reliable. When there is good reason to expect higher accuracy (lower variance) in some b_i, those equations should be weighted more heavily. *With what weights* $w_1, \cdots, w_m$? And if the b_i are not independent, a ***covariance matrix*** Σ gives the statistics of the errors. Here are key topics in this section:

1. Weighted least squares and $A^{\mathrm{T}}CA\widehat{x} = A^{\mathrm{T}}C\boldsymbol{b}$
2. Variances $\sigma_1^2, \ldots, \sigma_m^2$ and the covariance matrix Σ
3. Important probability distributions: binomial, Poisson, and normal
4. Principal Component Analysis (PCA) to find combinations with greatest variance.

Weighted Least Squares

To include weights in the m equations $Ax = \boldsymbol{b}$, multiply each equation i by a weight w_i. Put those m weights into a diagonal matrix W. *We are replacing* $Ax = \boldsymbol{b}$ *by* $WAx = W\boldsymbol{b}$. The equations are no more and no less solvable—we expect to use least squares.

The least squares equation $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}\boldsymbol{b}$ changes to $(WA)^{\mathrm{T}}WA\widehat{x} = (WA)^{\mathrm{T}}W\boldsymbol{b}$. The matrix $C = W^{\mathrm{T}}W$ is inside $(WA)^{\mathrm{T}}WA$, in the middle of weighted least squares.

Weighted least squares ***$C = W^{\mathrm{T}}W$ is in the n equations for $\widehat{x}$*** $\quad A^{\mathrm{T}}CA\widehat{x} = A^{\mathrm{T}}C\boldsymbol{b}$ (1)

When $n = 1$ and A = column of 1's, $\widehat{x}$ changes from an average to a weighted average:

$$\textbf{Simplest case} \quad \widehat{x} = \frac{b_1 + \cdots + b_m}{m} \quad \text{changes to} \quad \widehat{x}_W = \frac{w_1^2 b_1 + \cdots + w_m^2 b_m}{w_1^2 + \cdots + w_m^2}. \qquad (2)$$

This average $\widehat{x}_W$ gives greatest weight to the observations b_i that have the largest w_i. We always assume that errors have *zero mean.* (Subtract the mean if necessary, so there is no one-sided bias in the measurements.)

How should we choose the weights w_i? This depends on the reliability of b_i. If that observation has variance σ_i^2, then the root mean square error in b_i is σ_i. **When we divide the equations by** $\sigma_1, \ldots \sigma_m$ (left side together with right side), **all variances will equal 1.** So the weight is $w_i = 1/\sigma_i$ and the diagonal of $C = W^{\mathrm{T}}W$ contains the numbers $1/\sigma_i^2$.

The statistically correct matrix is $C = \operatorname{diag}(1/\sigma_1^2, \ldots, 1/\sigma_m^2)$.

This is correct provided the errors e_i and e_j in different equations are statistically independent. If the errors are dependent, off-diagonal entries show up in the covariance matrix Σ. The good choice is still $C = \Sigma^{-1}$ as described in this section.

Mean and Variance

The two crucial numbers for a random variable are its **mean** m and its **variance** σ^2. The "expected value" E[e] is found from the probabilities $p_1, p_2, \ldots$ of the possible errors $e_1, e_2, \ldots$ (and the variance σ^2 is always measured around the mean).

For a discrete random variable, the error e_j has probability p_j (the p_j add to 1):

$$\textbf{Mean } m = \mathrm{E}[e] = \sum e_j p_j \quad \textbf{Variance } \sigma^2 = \mathrm{E}[(e-m)^2] = \sum (e_j - m)^2 p_j \qquad (3)$$

Example 1 Flip a fair coin. The result is 1 (for heads) or 0 (for tails). Those events have equal probabilities $p_0 = p_1 = 1/2$. The mean is $m = 1/2$ and the variance is $\sigma^2 = 1/4$:

$$\text{Mean} = (0)\frac{1}{2} + (1)\frac{1}{2} \qquad \text{Variance} = \left(0 - \frac{1}{2}\right)^2 \frac{1}{2} + \left(1 - \frac{1}{2}\right)^2 \frac{1}{2} = \frac{1}{4}.$$

Example 2 **(Binomial)** Flip the fair coin N times and count heads. With 3 flips, we see $M = 0, 1, 2$, or 3 heads. The chances are $1/8, 3/8, 3/8, 1/8$. There are three ways to see $M = 2$ heads: HHT, HTH, and THH, and only HHH for $M = 3$ heads.

For all N, the number of ways to see M heads is the binomial coefficient "N *choose* M". Divide by the total number 2^N of all possible outcomes to get the probability for each M:

$$\begin{array}{l} \boldsymbol{M} \textbf{ heads in} \\ \boldsymbol{N} \textbf{ coin flips} \end{array} \quad p_M = \frac{1}{2^N}\binom{N}{M} = \frac{1}{2^N}\frac{N!}{M!(N-M)!} \qquad \text{Check } \frac{1}{2^3}\frac{3!}{2!\,1!} = \frac{3}{8} \qquad (4)$$

Gamblers know this instinctively. The probabilities p_M add to $\left(\frac{1}{2} + \frac{1}{2}\right)^N = 1$. The mean value of the number of heads is $m = N/2$. The variance around m turns out to be $\sigma^2 = N/4$. The standard deviation $\sigma = \sqrt{N}/2$ measures the expected spread around the mean.

Example 3 **(Poisson)** A very unfair coin (small $p << \frac{1}{2}$) is flipped very often (large N). **The product** $\lambda = pN$ **is kept fixed**. The high probability of tails is $1 - p$ each time. So the chance p_0 of no heads in N flips (tails every time) is $(1-p)^N = (1 - \lambda/N)^N$. For large N this approaches $e^{-\lambda}$. The probability p_j of j heads in N very unfair flips comes out neatly in terms of the crucial number $\lambda = pN$:

$$\textbf{Poisson probabilities} \quad p_j = \frac{\lambda^j}{j!}e^{-\lambda} \quad \textbf{Mean } m = \lambda \quad \textbf{Variance } \sigma^2 = \lambda \qquad (5)$$

Poisson applies to counting infrequent events (low p) over a long time T. Then $\lambda = pT$.

A *continuous* random variable will have a probability *density* function $p(x)$ instead of $p_1, p_2, \ldots$. "An outcome between x and $x + dx$ has probability $p(x)\,dx$." The total probability is $\int p(x)\,dx = 1$, since some outcome must happen. Sums become integrals:

$$\textbf{Mean } m = \textbf{Expected value} = \int xp(x)\,dx \quad \textbf{Variance } \sigma^2 = \int (x-m)^2 p(x)\,dx. \quad (6)$$

The outstanding example of a probability density function $p(x)$ (called the pdf) is the ***normal distribution*** $\mathrm{N}(0, \sigma)$. This has mean zero by symmetry. Its variance is σ^2:

$$\textbf{Normal (Gaussian)} \qquad p(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-x^2/2\sigma^2} \quad \text{with} \quad \int_{-\infty}^{\infty} p(x)\,dx = 1. \qquad (7)$$

The graph of $p(x)$ is the famous bell-shaped curve. The integral of $p(x)$ from $-\sigma$ to σ is the probability that a random sample is less than one standard deviation σ from the mean. This is near $2/3$. MATLAB's **randn** uses the normal distribution with $\sigma = 1$.

This normal $p(x)$ appears everywhere because of the ***Central Limit Theorem***: The average over many independent trials of another distribution (like binomial) will approach a normal distribution as $N \to \infty$. A shift produces $m = 0$ and rescaling produces $\sigma = 1$.

$$\textbf{Normalized headcount} \qquad x = \frac{M - \text{mean}}{\sigma} = \frac{M - N/2}{\sqrt{N}/2} \longrightarrow \text{Normal } \mathrm{N}(0, 1).$$

The Covariance Matrix

Now run m different experiments at once. They might be independent, or there might be some correlation between them. Each measurement $\boldsymbol{b}$ is now a *vector* with m components. Those components are the outputs b_i from the m experiments.

If we measure distances from the means m_i, each error $e_i = b_i - m_i$ has *mean zero*. If two errors e_i and e_j are *independent* (no relation between them), their product $e_i e_j$ also has mean zero. But if the measurements are by the same observer at nearly the same time, the errors e_i and e_j could tend to have the same sign or the same size. ***The errors in the m experiments could be correlated.*** The products $e_i e_j$ are weighted by p_{ij} (their probability): *covariance* $\sigma_{ij} = \sum\sum p_{ij} e_i e_j$. The sum of $e_i^2 p_{ii}$ is the variance σ_i^2:

$$\textbf{Covariance} \qquad \sigma_{ij} = \sigma_{ji} = \mathrm{E}[e_i e_j] = \textbf{expected value of } (e_i \textbf{ times } e_j). \qquad (8)$$

This is the (i, j) and (j, i) entry of the ***covariance matrix*** Σ. The (i, i) entry is $\sigma_{ii} = \sigma_i^2$.

Example 4 **(Multivariate normal)** For m random variables, the probability density function moves from $p(x)$ to $p(\boldsymbol{b}) = p(b_1, \ldots, b_m)$. The normal distribution with mean zero was controlled by one positive number σ^2. Now $p(\boldsymbol{b})$ is controlled by an m by m positive definite matrix Σ. This is the covariance matrix and its determinant is $|\Sigma|$:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} \quad \text{becomes} \quad p(\boldsymbol{b}) = \frac{1}{(2\pi)^{m/2}|\Sigma|^{1/2}} e^{-\boldsymbol{b}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{b}/2}$$

The integral of $p(\boldsymbol{b})$ over m-dimensional space is 1. The integral of $\boldsymbol{b}\boldsymbol{b}^{\mathrm{T}} p(\boldsymbol{b})$ is Σ.

The good way to handle that exponent $-\boldsymbol{b}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{b}/2$ is to use the eigenvalues and orthonormal eigenvectors of Σ (*linear algebra enters here*). When $\Sigma = Q\Lambda Q^{\mathrm{T}} = Q\Lambda Q^{-1}$, replacing $\boldsymbol{b}$ by $Q\boldsymbol{c}$ will split $p(\boldsymbol{b})$ into m one-dimensional normal distributions:

$$\exp\left(-\boldsymbol{b}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{b}/2\right) = \exp\left(-\boldsymbol{c}^{\mathrm{T}}\Lambda^{-1}\boldsymbol{c}/2\right) = \left(e^{-c_1^2/2\lambda_1}\right)\cdots\left(e^{-c_m^2/2\lambda_m}\right).$$

The determinant has $|\Sigma|^{1/2} = |\Lambda|^{1/2} = (\lambda_1\cdots\lambda_m)^{1/2}$. Each integral over $-\infty < c_i < \infty$ is back to one dimension, where $\lambda = \sigma^2$. Notice the wonderful fact that after any linear transformation (here $\boldsymbol{c} = Q^{-1}\boldsymbol{b}$), we still have a multivariate normal distribution.

We could even reach variances $= 1$ by including $\sqrt{\Lambda}$ in the change from $\boldsymbol{b}$ to z:

Standard normal $\quad \boldsymbol{b} = \sqrt{\Lambda}Qz$ changes $p(\boldsymbol{b})d\boldsymbol{b}$ to $p(z)dz = \dfrac{e^{-z^{\mathrm{T}}z/2}}{(2\pi)^{m/2}}dz$

This tells us the right weight matrix W to bring $A\boldsymbol{x} = \boldsymbol{b}$ back to ordinary least squares for $WA\boldsymbol{x} = W\boldsymbol{b}$. We want $W\boldsymbol{b}$ to become the standard normal z. So W will be the inverse of $\sqrt{\Lambda}Q$. Better than that, ***$C = W^{\mathrm{T}}W$ is the inverse of $Q\Lambda Q^{\mathrm{T}}$ which is Σ.***

Summary For independent errors, Σ is the diagonal matrix $\mathbf{diag}(\sigma_1^2,\ldots,\sigma_m^2)$. This is the usual choice. The right weights w_i for the equations $A\boldsymbol{x} = \boldsymbol{b}$ are $1/\sigma_1,\ldots,1/\sigma_m$ (this will equalize all variances to 1). The right matrix $C = W^{\mathrm{T}}W$ in the middle of the weighted least squares equations is exactly Σ^{-1}:

Weighted least squares $\quad A^{\mathrm{T}}\Sigma^{-1}A\widehat{x} = A^{\mathrm{T}}\Sigma^{-1}\boldsymbol{b} \quad (9)$

This choice of weighting returns $A\boldsymbol{x} = \boldsymbol{b}$ to a least squares problem $WA\boldsymbol{x} = W\boldsymbol{b}$ with equally reliable and independent errors. The usual equation $(WA)^{\mathrm{T}}WA\widehat{x} = (WA)^{\mathrm{T}}W\boldsymbol{b}$ is the same as (9).

It was Gauss who found this *best linear unbiased estimate* $\widehat{x}$. Unbiased because the mean of $\boldsymbol{x} - \widehat{x}$ is zero, linear because of equation (9), best because the covariance of $\boldsymbol{x} - \widehat{x}$ is as small as possible. That covariance (for error in $\widehat{x}$, not error in $\boldsymbol{b}$!) is important:

Covariance of the best $\widehat{x}$ $\quad P = \mathrm{E}\left[(x-\widehat{x})(x-\widehat{x})^{\mathrm{T}}\right] = \left(A^{\mathrm{T}}\Sigma^{-1}A\right)^{-1}. \quad (10)$

Example 5 Your pulse rate is measured ten times by independent doctors, all equally reliable. The mean error of each b_i is zero, and each variance is σ^2. Then $\Sigma = \sigma^2 I$. The ten equations $x = b_i$ produce the 10 by 1 matrix A of all ones. The best estimate $\widehat{x}$ is the average of the ten b_i. *The variance of that average value $\widehat{x}$ is the number P*:

$$P = (A^{\mathrm{T}}\Sigma^{-1}A)^{-1} = \sigma^2/10 \quad \text{so averaging reduces the variance.}$$

This matrix $P = (A^{\mathrm{T}}\Sigma^{-1}A)^{-1}$ tells how reliable is the result $\widehat{x}$ of the experiment (Problem 6). P does not depend on the b's in the actual experiment! Those b's have probability distributions. Each experiment produces a sample value of $\widehat{x}$ from a sample $\boldsymbol{b}$.

When a small Σ gives good reliability of the inputs $\boldsymbol{b}$, a small P gives good reliability of the outputs $\widehat{x}$. The key formula $P = (A^{\mathrm{T}}\Sigma^{-1}A)^{-1}$ connects those covariances.

Principal Component Analysis

These paragraphs are about finding useful information in a data matrix A. Start by measuring m properties (m features) of n samples. These could be grades in m courses for n students (a row for each course, a column for each student). From each row, subtract its average so the sample means are zero. We look for a ***combination of courses*** and/or ***combination of students*** for which the data provides the most information.

Information is "distance from randomness" and it is measured by **variance**. A large variance in course grades means greater information than a small variance.

The key matrix idea is the Singular Value Decomposition $A = U\Sigma V^{\mathrm{T}}$. We are back again to $A^{\mathrm{T}}A$ and AA^{T}, because their unit eigenvectors are the singular vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ in V and $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ in U. The singular values in the diagonal matrix Σ (not the covariance) are in decreasing order and σ_1 is the most important. Weighting the m courses by the components of $\boldsymbol{u}_1$ gives a "***master course***" or "***eigencourse***" with the most significant grades.

Example 6 Suppose the grades A, B, C, F are worth **4, 2, 0, −6** points. If each course and each student has one of each grade, then all means are zero. Here is the grade matrix A with $(1,1,1,1)$ in its nullspace (rank 3). To keep integers, the SVD of A will be written as $2U$ times $\Sigma/4$ times $(2V)^{\mathrm{T}}$. So the σ's are $12, 8, 4$:

$$\begin{bmatrix} -6 & 2 & 0 & 4 \\ 0 & 4 & -6 & 2 \\ 4 & 0 & 2 & -6 \\ 2 & -6 & 4 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 1 & -1 \\ -1 & -1 & 1 \\ 1 & -1 & -1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & & \\ & 2 & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 & -1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

Weighting the rows (the courses) by $\boldsymbol{u}_1 = \frac{1}{2}(-1,-1,1,1)$ will give the *eigencourse*. Weighting the columns (the students) by $\boldsymbol{v}_1 = \frac{1}{2}(1,-1,1,-1)$ gives the *eigenstudent*. The fraction of the grade matrix that is "explained" by that one course and student is $\sigma_1^2/(\sigma_1^2+\sigma_2^2+\sigma_3^2) = 9/14$. The σ's in the SVD are the variances σ^2.

I guess this master course is what a Director of Admissions is looking for. If all grades in gym are the same, that row of A will be all zero—and gym is not part of the master course. Probably calculus is a part, but what about students who don't take calculus? The problem of ***missing data*** (holes in the matrix A) is extremely difficult for social sciences and the census and so much of the statistics of experiments.

Gene expression data Determining the functions of genes, and combinations of genes, is a central problem of genetics. Which genes combine to give which properties? Which genes malfunction to give which diseases?

We now have an incredibly fast way to find gene expression data in the lab. A gene microarray is often packed onto an Affymetrix chip, measuring tens of thousands of genes from one sample (one person). The understanding of genetic data (*bioinformatics*) has become a tremendous application of linear algebra.

Problem Set 8.6

1 Which line $Ct + D$ is the best fit to the three independent measurements $1, 2, 4$ at times $t = 0, 1, 2$ if the variances $\sigma_1^2, \sigma_2^2, \sigma_3^2$ are $1, 1, 2$? Use weights $w_i = 1/\sigma_i$.

2 In Problem 1, suppose that the third measurement is **totally unreliable**. The variance σ_3^2 becomes infinite. Then the best line will not use _____. Find the line that goes through the first two points and solves the first two equations in $Ax = \boldsymbol{b}$ exactly.

3 In Problem 1, suppose that the third measurement is **totally reliable**. The variance σ_3^2 approaches zero. Now the best line will go through the third point exactly. Choose that line to minimize the sum of squares of the first two errors.

4 A single flip of a fair coin (0 or 1) has mean $m = 1/2$ and variance $\sigma^2 = 1/4$. This was Example 1. For the sum of two flips, the mean is $m = 1$. Compute the variance σ^2 around this mean, using the outcomes $0, 1, 2$ with their probabilities.

5 Instead of adding the flip results, make them two independent experiments. The outcome is $(0, 0), (1, 0), (0, 1)$ or $(1, 1)$. What is the covariance matrix Σ?

6 Change Example 1 so that the coin flip can be unfair. The probability is p for heads and $1 - p$ for tails. Find the mean m and the variance σ^2 of this distribution.

7 For two independent measurements $x = b_1$ and $x = b_2$, the best $\widehat{x}$ should be some weighted average $\widehat{x} = ab_1 + (1 - a)b_2$. When b_1 and b_2 have mean zero and variances σ_1^2 and σ_2^2, the variance of $\widehat{x}$ will be $P = a^2\sigma_1^2 + (1 - a)^2\sigma_2^2$. ***Choose the number a that minimizes P***: $dP/da = 0$.

Show that this a gives the $\widehat{x}$ in equation (2) which the text claimed is best, using weights $w_1 = 1/\sigma_1$ and $w_2 = 1/\sigma_2$.

8 The least squares estimate correctly weighted by Σ^{-1} is $\widehat{x} = (A^{\mathrm{T}}\Sigma^{-1}A)^{-1}A^{\mathrm{T}}\Sigma^{-1}\boldsymbol{b}$. ***Call that*** $\widehat{x} = L\boldsymbol{b}$. If $\boldsymbol{b}$ contains an error vector $\boldsymbol{e}$, then $\widehat{x}$ contains the error $L\boldsymbol{e}$.

The covariance matrix of those output errors $L\boldsymbol{e}$ is their expected value (average value) $P = \mathrm{E}\left[(L\boldsymbol{e})(L\boldsymbol{e})^{\mathrm{T}}\right] = L\mathrm{E}\left[\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}\right]L^{\mathrm{T}} = L\Sigma L^{\mathrm{T}}$. **Problem**: Do the multiplication $L\Sigma L^{\mathrm{T}}$ to show that P equals $(A^{\mathrm{T}}\Sigma^{-1}A)^{-1}$ as predicted in equation (10).

9 Change the grades to $3, 1, -1, -3$ for A, B, C, F. Show that the SVD of this grade matrix has the same $\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{v}_1, \boldsymbol{v}_2$ (same eigencourses) as in Example 5, but now A has rank 2.

Grade matrix $$A = \begin{bmatrix} 3 & -1 & 1 & -3 \\ -1 & 3 & -3 & 1 \\ -3 & 1 & -1 & 3 \\ 1 & -3 & 3 & -1 \end{bmatrix}$$

Notes One way to deal with missing entries in A is to complete the matrix to have minimum rank. And statistics makes major use of the pseudoinverse A^+ (which is exactly the left inverse $(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ from the normal equation when $A^{\mathrm{T}}A$ is invertible).

8.7 Computer Graphics

Computer graphics deals with images. The images are moved around. Their scale is changed. Three dimensions are projected onto two dimensions. All the main operations are done by matrices—but the shape of these matrices is surprising.

The transformations of three-dimensional space are done with 4 *by* 4 *matrices.* You would expect 3 by 3. The reason for the change is that one of the four key operations cannot be done with a 3 by 3 matrix multiplication. Here are the four operations:

Translation	**(shift the origin to another point $P_0 = (x_0, y_0, z_0)$)**
Rescaling	**(by c in all directions or by different factors c_1, c_2, c_3)**
Rotation	**(around an axis through the origin or an axis through P_0)**
Projection	**(onto a plane through the origin or a plane through P_0).**

Translation is the easiest—just add (x_0, y_0, z_0) to every point. But this is not linear! No 3 by 3 matrix can move the origin. So we change the coordinates of the origin to $(0, 0, 0, 1)$. This is why the matrices are 4 by 4. The "*homogeneous coordinates*" of the point (x, y, z) are $(x, y, z, 1)$ and we now show how they work.

1. Translation Shift the whole three-dimensional space along the vector v_0. The origin moves to (x_0, y_0, z_0). This vector v_0 is added to every point v in $\mathbf{R}^3$. Using homogeneous coordinates, the 4 by 4 matrix T shifts the whole space by v_0:

$$\textbf{\textit{Translation matrix}} \qquad T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ x_0 & y_0 & z_0 & 1 \end{bmatrix}.$$

Important: *Computer graphics works with row vectors.* We have row times matrix instead of matrix times column. You can quickly check that $[0 \ \ 0 \ \ 0 \ \ 1]\,T = [x_0 \ \ y_0 \ \ z_0 \ \ 1]$.

To move the points $(0, 0, 0)$ and (x, y, z) by v_0, change to homogeneous coordinates $(0, 0, 0, 1)$ and $(x, y, z, 1)$. Then multiply by T. A row vector times T gives a row vector. ***Every v moves to $v + v_0$***: $[x \ \ y \ \ z \ \ 1]\,T = [x + x_0 \ \ y + y_0 \ \ z + z_0 \ \ 1]$.

The output tells where any v will move. (It goes to $v + v_0$.) Translation is now achieved by a matrix, which was impossible in $\mathbf{R}^3$.

2. Scaling To make a picture fit a page, we change its width and height. A Xerox copier will rescale a figure by 90%. In linear algebra, we multiply by .9 times the identity matrix. That matrix is normally 2 by 2 for a plane and 3 by 3 for a solid. In computer graphics, with homogeneous coordinates, the matrix is *one size larger:*

$$\textbf{\textit{Rescale the plane:}} \quad S = \begin{bmatrix} .9 & & \\ & .9 & \\ & & 1 \end{bmatrix} \qquad \textbf{\textit{Rescale a solid:}} \quad S = \begin{bmatrix} c & 0 & 0 & 0 \\ 0 & c & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Important: S is not cI. We keep the "1" in the lower corner. Then $[x, y, 1]$ times S is the correct answer in homogeneous coordinates. The origin stays in its normal position because $[0\ 0\ 1]S = [0\ 0\ 1]$.

If we change that 1 to c, the result is strange. ***The point* (cx, cy, cz, c) *is the same as* $(x, y, z, 1)$.** The special property of homogeneous coordinates is that *multiplying by cI does not move the point.* The origin in $\mathbf{R}^3$ has homogeneous coordinates $(0, 0, 0, 1)$ and $(0, 0, 0, c)$ for every nonzero c. This is the idea behind the word "homogeneous."

Scaling can be different in different directions. To fit a full-page picture onto a half-page, scale the y direction by $\frac{1}{2}$. To create a margin, scale the x direction by $\frac{3}{4}$. The graphics matrix is diagonal but not 2 by 2. It is 3 by 3 to rescale a plane and 4 by 4 to rescale a space:

$$\textbf{\textit{Scaling matrices}} \quad S = \begin{bmatrix} \frac{3}{4} & & \\ & \frac{1}{2} & \\ & & 1 \end{bmatrix} \quad \text{and} \quad S = \begin{bmatrix} c_1 & & & \\ & c_2 & & \\ & & c_3 & \\ & & & 1 \end{bmatrix}.$$

That last matrix S rescales the x, y, z directions by positive numbers c_1, c_2, c_3. The extra column in all these matrices leaves the extra 1 at the end of every vector.

Summary The scaling matrix S is the same size as the translation matrix T. They can be multiplied. To translate and then rescale, multiply $\boldsymbol{v}TS$. To rescale and then translate, multiply $\boldsymbol{v}ST$. Are those different? *Yes.*

The point (x, y, z) in $\mathbf{R}^3$ has homogeneous coordinates $(x, y, z, 1)$ in $\mathbf{P}^3$. This "projective space" is not the same as $\mathbf{R}^4$. It is still three-dimensional. To achieve such a thing, (cx, cy, cz, c) is the same point as $(x, y, z, 1)$. Those points of projective space $\mathbf{P}^3$ are really lines through the origin in $\mathbf{R}^4$.

Computer graphics uses ***affine*** transformations, *linear plus shift.* An affine transformation T is executed on $\mathbf{P}^3$ by a 4 by 4 matrix with a special fourth column:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 \\ a_{21} & a_{22} & a_{23} & 0 \\ a_{31} & a_{32} & a_{33} & 0 \\ a_{41} & a_{42} & a_{43} & 1 \end{bmatrix} = \begin{bmatrix} T(1,0,0) & 0 \\ T(0,1,0) & 0 \\ T(0,0,1) & 0 \\ T(0,0,0) & 1 \end{bmatrix}.$$

The usual 3 by 3 matrix tells us three outputs, this tells four. The usual outputs come from the inputs $(1, 0, 0)$ and $(0, 1, 0)$ and $(0, 0, 1)$. When the transformation is linear, three outputs reveal everything. When the transformation is affine, the matrix also contains the output from $(0, 0, 0)$. Then we know the shift.

3. Rotation A rotation in $\mathbf{R}^2$ or $\mathbf{R}^3$ is achieved by an orthogonal matrix Q. The determinant is $+1$. (With determinant -1 we get an extra reflection through a mirror.) Include the extra column when you use homogeneous coordinates!

$$\textbf{\textit{Plane rotation}} \quad Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad \text{becomes} \quad R = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

This matrix rotates the plane around the origin. ***How would we rotate around a different point*** $(4,5)$? The answer brings out the beauty of homogeneous coordinates. ***Translate*** $(4,5)$ ***to*** $(0,0)$***, then rotate by*** θ***, then translate*** $(0,0)$ ***back to*** $(4,5)$:

$$vT_-RT_+ = \begin{bmatrix} x & y & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & -5 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 4 & 5 & 1 \end{bmatrix}.$$

I won't multiply. The point is to apply the matrices one at a time: v translates to vT_-, then rotates to vT_-R, and translates back to vT_-RT_+. Because each point $\begin{bmatrix} x & y & 1 \end{bmatrix}$ is a row vector, T_- acts first. The center of rotation $(4,5)$—otherwise known as $(4,5,1)$—moves first to $(0,0,1)$. Rotation doesn't change it. Then T_+ moves it back to $(4,5,1)$. All as it should be. The point $(4,6,1)$ moves to $(0,1,1)$, then turns by θ and moves back.

In three dimensions, every rotation Q turns around an axis. The axis doesn't move—it is a line of eigenvectors with $\lambda = 1$. Suppose the axis is in the z direction. The 1 in Q is to leave the z axis alone, the extra 1 in R is to leave the origin alone:

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad R = \begin{bmatrix} & & & 0 \\ & Q & & 0 \\ & & & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Now suppose the rotation is around the unit vector $\boldsymbol{a} = (a_1, a_2, a_3)$. With this axis $\boldsymbol{a}$, the rotation matrix Q which fits into R has three parts:

$$Q = (\cos\theta)I + (1-\cos\theta)\begin{bmatrix} a_1^2 & a_1a_2 & a_1a_3 \\ a_1a_2 & a_2^2 & a_2a_3 \\ a_1a_3 & a_2a_3 & a_3^2 \end{bmatrix} - \sin\theta \begin{bmatrix} 0 & a_3 & -a_2 \\ -a_3 & 0 & a_1 \\ a_2 & -a_1 & 0 \end{bmatrix}. \quad (1)$$

The axis doesn't move because $\boldsymbol{a}Q = \boldsymbol{a}$. When $\boldsymbol{a} = (0,0,1)$ is in the z direction, this Q becomes the previous Q—for rotation around the z axis.

The linear transformation Q always goes in the upper left block of R. Below it we see zeros, because rotation leaves the origin in place. When those are not zeros, the transformation is affine and the origin moves.

4. Projection In a linear algebra course, most planes go through the origin. In real life, most don't. A plane through the origin is a vector space. The other planes are affine spaces, sometimes called "flats." An affine space is what comes from translating a vector space.

We want to project three-dimensional vectors onto planes. Start with a plane through the origin, whose unit normal vector is $\boldsymbol{n}$. (We will keep $\boldsymbol{n}$ as a column vector.) The vectors in the plane satisfy $\boldsymbol{n}^{\mathrm{T}}\boldsymbol{v} = 0$. ***The usual projection onto the plane is the matrix*** $\boldsymbol{I} - \boldsymbol{n}\boldsymbol{n}^{\mathrm{T}}$. To project a vector, multiply by this matrix. The vector $\boldsymbol{n}$ is projected to zero, and the in-plane vectors $\boldsymbol{v}$ are projected onto themselves:

$$(I - \boldsymbol{n}\boldsymbol{n}^{\mathrm{T}})\boldsymbol{n} = \boldsymbol{n} - \boldsymbol{n}(\boldsymbol{n}^{\mathrm{T}}\boldsymbol{n}) = \boldsymbol{0} \quad \text{and} \quad (I - \boldsymbol{n}\boldsymbol{n}^{\mathrm{T}})\boldsymbol{v} = \boldsymbol{v} - \boldsymbol{n}(\boldsymbol{n}^{\mathrm{T}}\boldsymbol{v}) = \boldsymbol{v}.$$

In homogeneous coordinates the projection matrix becomes 4 by 4 (but the origin doesn't move):

$$\textbf{\textit{Projection onto the plane}}\quad n^{\mathrm{T}}v = 0 \qquad P = \begin{bmatrix} & & & 0 \\ & I - nn^{\mathrm{T}} & & 0 \\ & & & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Now project onto a plane $n^{\mathrm{T}}(v - v_0) = 0$ that does *not* go through the origin. One point on the plane is v_0. This is an affine space (or a ***flat***). It is like the solutions to $Av = b$ when the right side is not zero. One particular solution v_0 is added to the nullspace—to produce a flat.

The projection onto the flat has three steps. Translate v_0 to the origin by T_-. Project along the n direction, and translate back along the row vector v_0:

$$\textbf{\textit{Projection onto a flat}}\qquad T_-PT_+ = \begin{bmatrix} I & 0 \\ -v_0 & 1 \end{bmatrix}\begin{bmatrix} I - nn^{\mathrm{T}} & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} I & 0 \\ v_0 & 1 \end{bmatrix}.$$

I can't help noticing that T_- and T_+ are inverse matrices: translate and translate back. They are like the elementary matrices of Chapter 2.

The exercises will include reflection matrices, also known as *mirror matrices.* These are the fifth type needed in computer graphics. A reflection moves each point twice as far as a projection—***the reflection goes through the plane and out the other side.*** So change the projection $I - nn^{\mathrm{T}}$ to $I - 2nn^{\mathrm{T}}$ for a mirror matrix.

The matrix P gave a "*parallel*" projection. All points move parallel to n, until they reach the plane. The other choice in computer graphics is a "*perspective*" projection. This is more popular because it includes foreshortening. With perspective, an object looks larger as it moves closer. Instead of staying parallel to n (and parallel to each other), the lines of projection come *toward the eye*—the center of projection. This is how we perceive depth in a two-dimensional photograph.

The basic problem of computer graphics starts with a scene and a viewing position. Ideally, the image on the screen is what the viewer would see. The simplest image assigns just one bit to every small picture element—called a ***pixel.*** It is light or dark. This gives a black and white picture with no shading. You would not approve. In practice, we assign shading levels between 0 and 2^8 for three colors like red, green, and blue. That means $8 \times 3 = 24$ bits for each pixel. Multiply by the number of pixels, and a lot of memory is needed!

Physically, a *raster frame buffer* directs the electron beam. It scans like a television set. The quality is controlled by the number of pixels and the number of bits per pixel. In this area, one standard text is *Computer Graphics: Principles and Practices* by Foley, Van Dam, Feiner, and Hughes (Addison-Wesley, 1995). The newer books still use homogeneous coordinates to handle translations. My best references were notes by Ronald Goldman and by Tony DeRose.

■ REVIEW OF THE KEY IDEAS ■

1. Computer graphics needs shift operations $T(\boldsymbol{v}) = \boldsymbol{v} + \boldsymbol{v}_0$ as well as linear operations $T(\boldsymbol{v}) = A\boldsymbol{v}$.

2. A shift in $\mathbf{R}^n$ can be executed by a matrix of order $n + 1$, using homogeneous coordinates.

3. The extra component 1 in $[x\ y\ z\ 1]$ is preserved when all matrices have the numbers $0, 0, 0, 1$ as last column.

Problem Set 8.7

1 A typical point in $\mathbf{R}^3$ is $x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}$. The coordinate vectors $\boldsymbol{i}$, $\boldsymbol{j}$, and $\boldsymbol{k}$ are $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$. The coordinates of the point are (x, y, z).

This point in computer graphics is $x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}$ + **origin**. Its homogeneous coordinates are (, , ,). Other coordinates for the same point are (, , ,).

2 A linear transformation T is determined when we know $T(\boldsymbol{i}), T(\boldsymbol{j}), T(\boldsymbol{k})$. For an affine transformation we also need $T($_____$)$. The input point $(x, y, z, 1)$ is transformed to $xT(\boldsymbol{i}) + yT(\boldsymbol{j}) + zT(\boldsymbol{k}) +$ _____.

3 Multiply the 4 by 4 matrix T for translation along $(1, 4, 3)$ and the matrix T_1 for translation along $(0, 2, 5)$. The product TT_1 is translation along _____.

4 Write down the 4 by 4 matrix S that scales by a constant c. Multiply ST and also TS, where T is translation by $(1, 4, 3)$. To blow up the picture around the center point $(1, 4, 3)$, would you use $\boldsymbol{v}ST$ or $\boldsymbol{v}TS$?

5 What scaling matrix S (in homogeneous coordinates, so 3 by 3) would produce a 1 by 1 square page from a standard 8.5 by 11 page?

6 What 4 by 4 matrix would move a corner of a cube to the origin and then multiply all lengths by 2? The corner of the cube is originally at $(1, 1, 2)$.

7 When the three matrices in equation 1 multiply the unit vector $\boldsymbol{a}$, show that they give $(\cos\theta)\boldsymbol{a}$ and $(1 - \cos\theta)\boldsymbol{a}$ and $\mathbf{0}$. Addition gives $\boldsymbol{a}Q = \boldsymbol{a}$ and the rotation axis is not moved.

8 If $\boldsymbol{b}$ is perpendicular to $\boldsymbol{a}$, multiply by the three matrices in 1 to get $(\cos\theta)\boldsymbol{b}$ and $\mathbf{0}$ and a vector perpendicular to $\boldsymbol{b}$. So $Q\boldsymbol{b}$ makes an angle θ with $\boldsymbol{b}$. ***This is rotation.***

9 What is the 3 by 3 projection matrix $I - \boldsymbol{n}\boldsymbol{n}^{\mathrm{T}}$ onto the plane $\frac{2}{3}x + \frac{2}{3}y + \frac{1}{3}z = 0$? In homogeneous coordinates add $0, 0, 0, 1$ as an extra row and column in P.

10 With the same 4 by 4 matrix P, multiply T_-PT_+ to find the projection matrix onto the plane $\frac{2}{3}x + \frac{2}{3}y + \frac{1}{3}z = 1$. The translation T_- moves a point on that plane (choose one) to $(0, 0, 0, 1)$. The inverse matrix T_+ moves it back.

11 Project $(3, 3, 3)$ onto those planes. Use P in Problem 9 and $T_- P T_+$ in Problem 10.

12 If you project a square onto a plane, what shape do you get?

13 If you project a cube onto a plane, what is the outline of the projection? Make the projection plane perpendicular to a diagonal of the cube.

14 The 3 by 3 mirror matrix that reflects through the plane $\boldsymbol{n}^{\mathrm{T}}\boldsymbol{v} = 0$ is $M = I - 2\boldsymbol{n}\boldsymbol{n}^{\mathrm{T}}$. Find the reflection of the point $(3, 3, 3)$ in the plane $\frac{2}{3}x + \frac{2}{3}y + \frac{1}{3}z = 0$.

15 Find the reflection of $(3, 3, 3)$ in the plane $\frac{2}{3}x + \frac{2}{3}y + \frac{1}{3}z = 1$. Take three steps $T_- M T_+$ using 4 by 4 matrices: translate by T_- so the plane goes through the origin, reflect the translated point $(3, 3, 3, 1)T_-$ in that plane, then translate back by T_+.

16 The vector between the origin $(0, 0, 0, 1)$ and the point $(x, y, z, 1)$ is the difference $\boldsymbol{v} =$ _____. In homogeneous coordinates, vectors end in _____. So we add a _____ to a point, not a point to a point.

17 If you multiply only the *last* coordinate of each point to get (x, y, z, c), you rescale the whole space by the number _____. This is because the point (x, y, z, c) is the same as $(\ , \ , \ , 1)$.