1. Outline 41 started an inquiry into the origin of *Sylvester's criterion*: a symmetric matrix is positive definite if and only if all determinants of square submatrices in its upper-left corner—its *principal minors*—are positive. The question came from a lecture in Prof. Ardila's Math 490 course. He gave a brief proof of the criterion, noted that it didn't reveal any intuitive approach to the result, and wondered how anyone ever discovered it. Ardila told me he'd found that proof in a not-too-old article in the *American Mathematical Monthly*. That information and Google led class members to the reference Gilbert 1991. Unfortunately, Gilbert had given no clue to the origin of the criterion. Digging in biographies of Sylvester proved fruitless, but Parshall's 1998 collection of his correspondence led us to a reference to a paper by Sylvester. That search was interrupted, but Google led me later to the required volume 1 of Sylvester's 1904 collected works, online via a new service that I hadn't known about earlier. Unfortunately, that paper said nothing about the criterion. However, nearby was a much more pertinent paper: Sylvester 1852.
2. Sylvester concluded that paper with the statement and proof of a landmark in the theory of quadratic forms: his *Law of Inertia*. His arguments explicitly involved eigenvalues of the corresponding symmetric matrices. In one hand-waving episode, he also referred to the determinants of *all* the submatrices, not just the principal minors. In a footnote at the beginning of the paper, he wrote that Jacobi and Borchardt had given comparable discussions, but Borchardt's involved Sturm's functions. Borchardt's approach doesn't seem promising for this inquiry.
3. I consulted Klein 1926–1927, a great source of historical information. (It has been translated as Klein [1926–1927] 1979, but I have only the original German edition with me.) In volume 2, 16–21, Klein discussed this question without referring directly to Sylvester's criterion. He sketched Jacobi's presentation, in enough detail to give the gist, and cited Jacobi 1857, which is online. Following Jacobi's presentation would require prior absorption of algebra that is not familiar to me or many others these days. But I can present the argument in full for the three-dimensional case; it's quite intuitive.
4. For each positive integer n, an n-ary (homogeneous) *quadratic form* is a function f defined for all scalars $x_1, \dots, x_n$ by an equation

   $$f(x_1, \dots, x_n) = \sum_{i,j=1}^{n} a_{i,j} x_i x_j$$

   for some $n \times n$ matrix A of scalars $a_{i,j}$. The symmetric $n \times n$ matrix B of scalars $b_{i,j} = ½(a_{i,j} + a_{j,i})$ satisfies the analogous equation

   $$f(x_1, \dots, x_n) = \sum_{i,j=1}^{n} b_{i,j} x_i x_j$$

   for all $x_1, \dots, x_n$. This discussion is really about forms, not matrices, hence there's no loss in generality in always assuming that the matrix A that defines a quadratic form is symmetric.

5. Let $n = 3$, for example, so that

$$f(x_1,x_2,x_3) = a_{1,1}x_1^{\,2} + 2a_{1,2}x_1x_2 + 2a_{1,3}x_1x_3 + a_{2,2}x_2^{\,2} + 2a_{2,3}x_2x_3 + a_{3,3}x_3^{\,2}.$$

Assume that the principal minors

$$P_1 = a_{1,1}, \qquad P_2 = \det\begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} = a_{1,1}a_{2,2} - a_{1,2}^2, \qquad P_3 = \det A$$

are nonzero. Subtracting from $f(x_1,x_2,x_3)$ a scalar multiple of the square of any linear combination of x_1,x_2,x_3 yields another quadratic form f_1. Is it possible to do so in such a way that f_1 has fewer variables? If so, repetition of that process could result in a formula for $f(x_1,x_2,x_3)$ as a weighted sum of the squares of a linear combination of x_1,x_2,x_3, of a linear combination of x_2,x_3, and of x_3 itself. The weights would determine whether f always has positive values.

6. In fact, this first step is possible:

$$f(x_1,x_2,x_3) - \frac{(a_{1,1}x_1 + a_{1,2}x_2 + a_{1,3}x_3)^2}{a_{1,1}} = f_1(x_2,x_3),$$

where

$$f_1(x_2,x_3) = b_{2,2}x_2^{\,2} + 2b_{2,3}x_2x_3 + b_{3,3}x_3^{\,2}$$

$$b_{2,2} = a_{2,2} - \frac{a_{1,2}^2}{a_{1,1}} = \frac{P_2}{P_1} \neq 0$$

$$b_{2,3} = a_{2,3} - \frac{a_{1,2}a_{1,3}}{a_{1,1}} = \frac{\det\begin{bmatrix} a_{1,1} & a_{1,3} \\ a_{2,1} & a_{2,3} \end{bmatrix}}{P_1} = b_{3,2}$$

$$b_{3,3} = a_{3,3} - \frac{a_{1,3}^2}{a_{1,1}} = \frac{\det\begin{bmatrix} a_{1,1} & a_{1,3} \\ a_{3,1} & a_{3,3} \end{bmatrix}}{P_1}.$$

Now repeat this process with f_1 in place of f:

$$f_1(x_2,x_3) - \frac{(b_{2,2}x_2 + b_{2,3}x_3)^2}{b_{2,2}} = f_2(x_3),$$

where

$$f_2(x_3) = c_{3,3}x_3^{\,2}$$

$$c_{3,3} = b_{3,3} - \frac{b_{2,3}^2}{b_{2,2}} = \frac{b_{2,2}b_{3,3} - b_{2,3}^2}{b_{2,2}} = \frac{\det\begin{bmatrix} b_{2,2} & b_{2,3} \\ b_{2,3} & b_{3,3} \end{bmatrix}}{P_2 / P_1}$$

$$= \frac{P_1}{P_2}\det\begin{bmatrix} \frac{1}{P_1}\det\begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} & \frac{1}{P_1}\det\begin{bmatrix} a_{1,1} & a_{1,3} \\ a_{2,1} & a_{2,3} \end{bmatrix} \\ \frac{1}{P_1}\det\begin{bmatrix} a_{1,1} & a_{1,3} \\ a_{2,1} & a_{2,3} \end{bmatrix} & \frac{1}{P_1}\det\begin{bmatrix} a_{1,1} & a_{1,3} \\ a_{3,1} & a_{3,3} \end{bmatrix} \end{bmatrix}$$

$$= \frac{1}{P_1P_2}\det\begin{bmatrix} \det\begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} & \det\begin{bmatrix} a_{1,1} & a_{1,3} \\ a_{2,1} & a_{2,3} \end{bmatrix} \\ \det\begin{bmatrix} a_{1,1} & a_{1,3} \\ a_{2,1} & a_{2,3} \end{bmatrix} & \det\begin{bmatrix} a_{1,1} & a_{1,3} \\ a_{3,1} & a_{3,3} \end{bmatrix} \end{bmatrix}.$$

At this point, I think I'm proving case $n = 3$ of a theorem, whose general statement I don't know, about the determinants of $n \times n$ matrices and those of their 2×2 submatrices. The previous term

$$= \frac{1}{P_1P_2}\det\begin{bmatrix} a_{1,1}a_{2,2} - a_{1,2}^2 & a_{1,1}a_{2,3} - a_{1,2}a_{1,3} \\ a_{1,1}a_{2,3} - a_{1,2}a_{1,3} & a_{1,1}a_{3,3} - a_{1,3}^2 \end{bmatrix}$$

$$= ((a_{1,1}a_{2,2} - a_{1,2}^2)(a_{1,1}a_{3,3} - a_{1,3}^2) - (a_{1,1}a_{2,3} - a_{1,2}a_{1,3})^2)/(P_1P_2)$$

$$= a_{1,1}(a_{1,1}a_{2,2}a_{3,3} + 2a_{1,2}a_{1,3}a_{2,3} - a_{1,1}a_{2,3}^2 - a_{2,2}a_{1,3}^2 - a_{3,3}a_{1,2}^2)/(P_1P_2).$$

That is,

$$c_{3,3} = \det A / P_2 = P_3/P_2 \neq 0.$$

In summary,

$$f(x_1,x_2,x_3) = \frac{(a_{1,1}x_1 + a_{1,2}x_2 + a_{1,3}x_3)^2}{a_{1,1}} + \frac{(b_{2,2}x_2 + b_{2,3}x_3)^2}{b_{2,2}} + c_{3,3}{x_3}^2,$$

the desired formula. Setting $P_0 = 1$, substituting $a_{1,1} = P_1$, $b_{2,2} = P_2/P_1$, and $c_{3,3} = P_3/P_2$ yields Jacobi's and Klein's versions of this formula,

$$f(x_1,x_2,x_3) = \frac{(a_{1,1}x_1 + a_{1,2}x_2 + a_{1,3}x_3)^2}{P_0P_1} + \frac{(P_1b_{2,2}x_2 + P_1b_{2,3}x_3)^2}{P_1P_2} + \frac{\left(P_3x_3^2\right)}{P_2P_3},$$

$$= \frac{(a_{1,1}x_1 + a_{1,2}x_2 + a_{1,3}x_3)^2}{P_1} + \frac{(P_1 b_{2,2}x_2 + P_1 b_{2,3}x_3)^2}{P_1 P_2} + \frac{\left(P_1 P_3 x_3^2\right)}{P_1 P_2 P_3}.$$

7. Jacobi and Klein claimed that their formulas generalize to apply to all positive integers n. Given a quadratic form f defined by a symmetric matrix A, Jacobi's formula shows how to define a matrix B of scalars $b_{i,j}$ so that for all $x_1,\ldots,x_n$,

$$f(x_1,\ldots,x_n) = \sum_{i=1}^{n} \frac{1}{P_{i-1}P_i} \left(\sum_{j=i}^{n} b_{i,j} x_j \right)^2 .$$

Klein's formula is similar, but has $P_1 \cdots P_i$ in place of $P_{i-1}P_i$. The only thing lacking for a proof of these is proper formulation of the theorem about determinants and minors referred to in the previous item, and formulation of a recursive proof that incorporates it.

8. Half of Sylvester's criterion is an immediate consequence of either formula: if the principal minors P_i for $i = 1,\ldots,n$ are all positive, then $f(x_1,\ldots,x_n) \geq 0$ for all $x_1,\ldots,x_n$. Moreover, if $f(x_1,\ldots,x_n) = 0$, then the last term in the outer sum of the formula must be zero, hence $x_n = 0$; this, and the fact that the penultimate term must be zero, implies $x_{n-1} = 0$, and so on. Thus $f(x_1,\ldots,x_n) > 0$ for all $x_1,\ldots,x_n$.
9. Internet searching revealed a paper, Johnson 1970, on this subject. Johnson proved that half of Sylvester's criterion by a different argument, and provided a rather clear argument for the converse.
10. Charles R. Johnson wrote that paper while an undergraduate at Northwestern University. It won first prize in an MAA contest for such papers, and could serve as a (rather short) *model for our master's expository papers*. Johnson earned the doctorate later from California Institute of Technology, and has long been a professor at the College of William and Mary. *MathSciNet* lists 338 papers by him!
11. It still remains to find the origin of the term *Sylvester's criterion*.

Postscript

12. Relaxing after posting the previous version of this epilog, I realized that the matrix

$$M = \begin{bmatrix} \frac{1}{P_1}\det\begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} & \frac{1}{P_1}\det\begin{bmatrix} a_{1,1} & a_{1,3} \\ a_{2,1} & a_{2,3} \end{bmatrix} \\ \frac{1}{P_1}\det\begin{bmatrix} a_{1,1} & a_{1,3} \\ a_{2,1} & a_{2,3} \end{bmatrix} & \frac{1}{P_1}\det\begin{bmatrix} a_{1,1} & a_{1,3} \\ a_{3,1} & a_{3,3} \end{bmatrix} \end{bmatrix}$$

near the top of page 3 is simply the intersection of all but the first rows and columns of the matrix that results from A by eliminating all but the first entry of the first column according to the standard Gauss procedure. No row interchanges are involved, so the procedure does not change the determinant: $\det A = a_{1,1} \det M = P_1 \det M$. Therefore

$$c_{3,3} = (P_1/P_2) \det M = (P_1 \det M)/P_2 = (\det A)/P_2 = P_3/P_2 \neq 0,$$

in accordance with the previous derivation.

13. This neat derivation generalizes to the $n \times n$ case: just regard $\det A$ as P_3, not the *largest* principal minor, but rather the *next* one after P_2. This observation allows easy construction of a recursive proof of the Jacobi or Klein formula.
14. I probably didn't notice the connection with Gauss elimination at first because when applied to symmetric matrices the indices can be switched; moreover, the 2×2 submatrices can be transposed without changing their determinants. I now realize also why this algebra appeared so familiar: I followed it about 28 years ago while preparing lectures for the CSc 810 analysis of algorithms course. Some familiar linear-algebra computations can be speeded up by implementing unfamiliar scalar-arithmetic operations in "firmware"—for example, implementing a "machine language" instruction that computes the determinant of a 2×2 matrix with scalar entries in specified registers, via an algorithm more efficient than the usual one that involves two multiplications and one subtraction.