Various Proofs of the Sylvester Criterion for Quadratic Forms

Giorgio Giorgi[1]

[1] Department of Economics and Management, Italy.

Correspondence: Giorgio Giorgi, Department of Economics and Management, Via S. Felice, 5 - 27100 Pavia, Italy.
E-mail: giorgio.giorgi@unipv.it

Abstract

In the first part of the paper we present several proofs of the so-called Sylvester criterion for quadratic forms; some of the said proofs are short and easy. In the second part of the paper we give an algebraic proof of the Sylvester criterion for quadratic forms subject to a linear homogeneous system.

Keywords

Quadratic forms, Sylvester criterion, positive definite quadratic forms, constrained quadratic forms.

AMS-2000 Math. Subj. Class.

1. Introduction

Given a (real) symmetric matrix A, of order n, and its associated quadratic form

$$Q(x) = x^\top A x, \quad x \in \mathbb{R}^n, \tag{1}$$

one of the most used and taught criteria to test the positive (or negative) definiteness of (1) is the so-called *Sylvester criterion.* Whereas the necessary part of the proof of this criterion is rather simple, the proof of the sufficient part is often omitted or is given in a somewhat long and intricate way. The aim of the first part of the present paper (Section 2) is to offer various proofs of the Sylvester criterion for quadratic forms, with the hope that some of the said proofs can be useful for a course of Linear Algebra delivered to undergraduates.

In the second part of the paper (Section 3) we give a relatively simple algebraic proof of a well-known necessary and sufficient determinantal criterion for positive (negative) definiteness of a quadratic form subject to a system of homogeneous linear equations. This determinantal criterion may therefore be considered a generalization of the Sylvester criterion to a "constrained quadratic form".

We recall that the i, j *minor* of the element a_{ij} of matrix A, of order n, denoted M^{ij}, is the determinant of the $(n-1) \times (n-1)$ matrix obtained by deleting the i-th row and the j-th column of A. The i, j *cofactor* of the element a_{ij} of A, denoted C^{ij}, is given by $(-1)^{i+j} M^{ij}$. We denote by $A_{(k)}$ the k-th order *leading principal submatrix* of A or k-th order *North-West principal submatrix* of A, i. e. the square submatrix of order k, consisting of the first k rows and first k columns of A, $k = 1, ..., n$. The k-th order *leading principal minor* or *North-West principal minor* of A is $\det(A_{(k)})$, denoted also Δ_k, $k = 1, ..., n$. The k-th order *principal minor* of the $n \times n$ matrix A, is the k-th order leading principal minor of $P^\top A P$, where P is some permutation matrix of order n. There are $n!/k!(n-k)!$ possible k-th order principal minors.

Theorem 1 (Sylvester criterion).   Let A be a real symmetric matrix of order n. Then (1) is positive definite (equivalently: matrix A is positive definite) if and only if

$$\Delta_1 > 0, \quad \Delta_2 > 0, ..., \Delta_k > 0, ..., \Delta_n > 0.$$

The quadratic form (1) is negative definite (equivalently: matrix A is negative definite) if and only if

$$\Delta_1 < 0, \quad \Delta_2 > 0, \quad \Delta_3 < 0, ..., (-1)^n \Delta_n > 0.$$

Remark 1.   We note that the second part of Theorem 1 follows at once from the first part, by remarking that if A is positive definite, then $-A$ is negative definite (if A is negative definite, then $-A$ is positive definite), and that, by basic properties of determinants, it holds $\det(-A_{(k)}) = (-1)^k \det(A_{(k)})$.

Other basic properties related to (1) are the following ones.

Theorem 2. Matrix A is positive definite if and only if all its eigenvalues are positive; A is negative definite if and only if all its eigenvalues are negative.

Contrary to the Sylvester criterion, the proof of Theorem 2 is quite short and simple, being based on the so-called "spectral theorem" or "principal axes theorem": any symmetric matrix can be diagonalized by an orthogonal transformation matrix.

Corollary 1. If A, symmetric, of order n, is positive definite, then

i) $a_{ii} > 0$, $i = 1, ..., n$.

ii) $\det(A) = \Delta_n > 0$.

Proof.

i)   Let us choose

$$x = [0, 0, ..., x_i, ..., 0]^\top, \quad x_i \neq 0.$$

Then $Q(x) = x^\top A x = a_{ii}(x_i)^2 > 0$, from which $a_{ii} > 0$.

(ii) Owing to Theorem 2, the eigenvalues of A are all positive and being $\det(A) = \prod\limits_{i=1}^{n} \lambda_i$, we have the thesis.   □

We note further that if A is positive definite, then obviously any North-West principal submatrix $A_{(k)}$ is positive definite, $k = 1, ..., n$.

Other useful properties concerning (1) are given by the following results (see, e. g., Horn and Johnson (1985)).

Theorem 3. The quadratic form (1) is positive definite if and only if there exists a nonsingular matrix Q of order n, not necessarily unique, such that $A = Q^\top Q$.

Theorem 4. The quadratic form (1) is positive definite if and only if $B^\top A B$ is positive definite for any $n \times m$ matrix B with $rank(B) = m$.

Corollary 2. If A is positive definite, then A^{-1} (which is a symmetric matrix, if A is symmetric) is positive definite too.

(It is sufficient to note that if $\lambda \neq 0$ is an eigenvalue of the nonsingular matrix A (if $\lambda = 0$, then A would be singular), then $1/\lambda$ is eigenvalue of A^{-1}).

2. Various Proofs of the Sylvester Criterion

A) A straightforward proof.

i) Proof of the necessary conditions.

The necessary conditions are usually proved in an easy way in most papers and books: for each k, $k = 1, ..., n$, let us choose a vector $x^k \in \mathbb{R}^n$, $x^k \neq 0$, with the first k components arbitrary real numbers and with the last $(n - k)$ components equal to zero:

$$x^k = [x_1, x_2, ..., x_k, 0, 0, ..., 0]^\top.$$

Let us define

$$z^k = [x_1, x_2, ..., x_k] \in \mathbb{R}^k.$$

It is easy to verify that

$$(z^k)^\top A_{(k)} z^k = x^\top A x > 0.$$

Therefore $A_{(k)}$ is positive definite. From Corollary 1 it follows that $\Delta_k = \det(A_{(k)}) > 0$, $k = 1, ..., n$.

ii) Proof of the sufficient conditions.

We use the "complete induction method". Obviously the condition is true for $k = 1$. In order to prove that the same is true for $k = 2$, we have to verify that

$$\{\Delta_1 = \det(a_{11}) > 0 \text{ and } \Delta_2 = \det(A_{(2)}) > 0\} \Longrightarrow A_{(2)} \text{ is positive definite}.$$

From $a_{11} > 0$ and $a_{11}a_{22} - (a_{12})^2 > 0$ it follows $a_{22} > 0$ and therefore that also $tr(A_{(2)}) = a_{11} + a_{22} > 0$. Being $\Delta_2 > 0$ and $tr(A_{(2)}) > 0$, taking the characteristic equation of $A_{(2)}$ into account, we get that the eigenvalues of $A_{(2)}$ are both positive and therefore $A_{(2)}$ is positive definite.

Then, we have to prove the implication:

$$\{A_{(k)} \text{ positive definite}, \Delta_{k+1} = \det(A_{(k+1)}) > 0\} \Longrightarrow A_{(k+1)} \text{ positive definite}.$$

Usually the above implication is proved by means of numerous algebraic transformations. It can be easily proved as follows. Let us suppose that $A_{(k+1)}$ is *not* positive definite. It follows that $A_{(k+1)}$ must posses at least two *negative* eigenvalues, say $\bar{\lambda}$ and $\hat{\lambda}$ (if $A_{(k+1)}$ would have only one negative eigenvalue, then $\Delta_{k+1} < 0$). As $A_{(k+1)}$ is symmetric, there exist two orthogonal eigenvectors, say x and y, corresponding to the said two negative eigenvalues:

$$x^\top y = 0.$$

We can therefore choose a linear combination of x and y, $u = \alpha x + \beta y$, such that $u \neq 0$ and the last entry of u is zero (x and y are both different from the zero vector, but they cannot have only the last component different from zero, otherwise they would not be orthogonal). Then we have

$$u^\top A_{(k+1)} u = \alpha^2 (x^\top A_{(k+1)} x) + \beta^2 (y^\top A_{(k+1)} y) < 0,$$

as from $A_{(k+1)} x = \bar{\lambda} x$, with $\bar{\lambda} < 0$, we get $x^\top A_{(k+1)} x = \bar{\lambda} x^\top x < 0$ and similarly we get $y^\top A_{(k+1)} y < 0$. But then $A_{(k)}$ would not be positive definite, against the induction assumption. So, all eigenvalues of $A_{(k+1)}$ must be positive, i. e. $A_{(k+1)}$ is positive definite. $\qquad\square$

B)   Another simplified proof of the sufficiency of the Sylvester criterion.

We have said that the proofs of *ii*) in the previous section **A)** are usually quite long and/or intricate. See, e. g., Bomze (1995), Mirsky (1955), Kemp and Kimura (1978), Simon and Blume (1994). An exception is given by Samuels (1966) whose proof we report here, for the reader's convenience. This proof relies on the following lemma.

Lemma 1.   Let

$$A = \left[\begin{array}{cc} A_{(n-1)} & a \\ a^\top & \beta \end{array} \right]$$

be an n-square symmetric matrix, with $A_{(n-1)}$ the (symmetric) North-West principal submatrix of A, of order $n-1$. Then, if $A_{(n-1)}$ is nonsingular, there exists a nonsingular matrix P such that

$$P^\top A P = \left[\begin{array}{cc} A_{(n-1)} & 0 \\ 0^\top & \beta \end{array} \right].$$

If, in addition, $\det(A_{(n-1)}) > 0$ and $\det(A) > 0$, then $\beta > 0$.

Proof.   Since $A_{(n-1)}$ is nonsingular, vector a has a unique representation as a linear combination of the columns of $A_{(n-1)}$. Let $\alpha_1, ..., \alpha_{n-1}$ be the coefficients of this linear combination of the said columns. Let $P^\top$ be the matrix obtained from the identity matrix I by replacing its last row by

$$[-\alpha_1, ..., -\alpha_{n-1}, 1].$$

Then P is nonsingulat and, by the symmetry of A, is the desired matrix. Since

$$\det(P^\top A P) = (\det(P))^2 \det(A) = \beta \det(A_{(n-1)}),$$

the second part of the lemma follows at once. $\qquad\square$

On the grounds of the previous lemma, the second "step" of the inductive proof of *ii*) sub A), is immediate: let A be an n-th square symmetric matrix with positive principal minors. Then A satisfies the assumptions of the lemma; the positive definiteness of $x^\top A x$ is equivalent to positive definiteness of

$$x^\top P^\top A P x = y^\top A_{(n-1)} y + (x_n)^2 \beta,$$

where y denotes the first $(n-1)$ components of $x : y = [x_1, x_2, ..., x_{n-1}]^\top$. This form is positive definite since the first term is positive by the induction assumption and the second term is positive, thanks to Lemma 1. $\qquad\square$

We point out that a similar and elegant proof of the above sufficient condition is provided by Hestenes (1975).

C)   Proofs by means of the Lagrange-Jacobi reduction formula.

There are various proofs of the Sylvester criterion which are based on the Lagrange-Jacobi reduction formula; some proofs are clear and efficient from a didactic point of view, but rather long, such as for example the proofs of Gantmacher (1959), Hadley (1961) and Hohn (1973). Some other proofs are more concise but also more difficult to grasp and therefore not too suitable for a course to undergraduates; it is the case, e. g., of the classical paper of Debreu (1952), of the books of Aleskerov and others (2011) and Murata (1977). Other proofs are incomplete, in the sense that they are performed for $n = 2$ or $n = 3$, such as in Bellman (1970). The proof of Truchon (1987) follows essentially the one of Debreu.

As the proof of the necessary conditions in the Sylvester criterion does not entail difficulties, we concentrate ourselves on the proof of the sufficiency part. We assume that all North-West principal minors of A are positive:

$$\Delta_1 > 0, \Delta_2 > 0, ..., \Delta_n > 0.$$

Let us consider the upper triangular matrix

$$B = \left[b_{ij}\right], \ i, j = 1, ..., n,$$

where the elements b_{ij} are defined as follows

$$b_{ij} = \begin{cases} 0, & \text{if } j < i \\ C_{(j)}^{ji}/\Delta_j, & \text{if } j \geqq i, \end{cases}$$

where $C_{(j)}^{ji}$ denotes the cofactor of a_{ji} in the North-West principal submatrix $A_{(j)}$, being, by definition, $C_{(1)}^{11} = 1$.

We note that matrix B is nonsingular, being

$$\det(B) = b_{11}b_{22}...b_{nn} = \frac{1}{\Delta_1} \frac{C_{(2)}^{22}}{\Delta_2} ... \frac{C_{(n)}^{nn}}{\Delta_n} =$$

$$= \frac{1}{\Delta_1} \frac{\Delta_1}{\Delta_2} ... \frac{\Delta_{n-1}}{\Delta_n} = \frac{1}{\Delta_n} = \frac{1}{\det(A)} \neq 0.$$

It follows that B has an inverse B^{-1}. We now introduce the vectors $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ and make the one-to-one transformation

$$x = By \Longleftrightarrow y = B^{-1}x.$$

From $Q(x) = x^\top A x$ we get

$$Q(y) = (By)^\top A By = y^\top B^\top A By = y^\top C y,$$

where $C = B^\top A B$. We note that $C = C^\top$, being $A = A^\top$.

Therefore $Q(y)$ is a symmetric quadratic form having the same sign of $Q(x)$. But matrix C is also a *diagonal matrix,* being lower triangular and symmetric. Moreover, it holds, with $\Delta_0 = 1$ by definition,

$$c_{jj} = b_{jj} = \Delta_{j-1}/\Delta_j, \ \ j = 1, 2, ..., n. \tag{2}$$

Indeed, in order to evaluate $C = B^\top A B$, it is convenient to evaluate first the product $AB = \left[\alpha_{ij}\right]$. As matrix B is upper triangular, we have

$$\alpha_{hj} = \sum_{k=1}^n a_{hk} b_{kj} = \left(\sum_{k=1}^j a_{hk} C_{(j)}^{jk}/\Delta_j\right).$$

The numerator in the last expression is equal to Δ_j, if $h = j$, it is equal to zero if $h < j$ (use the Laplace second theorem on determinants). Therefore $\alpha_{hj} = 0$ if $h < j$: hence matrix AB is lower triangular and we have $\alpha_{jj} = 1$, $j = 1, ..., n$. But $B^\top$ is a lower triangular matrix (it is the transpose of an upper triangular matrix) and therefore the product $C = B^\top(AB)$ will be lower triangular too. As $C = C^\top$, C is a diagonal matrix. Being $\alpha_{jj} = 1$, we have $c_{jj} = b_{jj}.1$, and taking into account the definition of b_{ij}, we get relation (2).

From

$$Q(y) = y^\top C y$$

we have the following Lagrange-Jacobi reduction formula, from which the Sylvester criterion is immediate:

$$Q(y) = \sum_{i=1}^n (\Delta_{i-1}/\Delta_i)(y_i)^2, \ \text{ with } \Delta_0 = 1. \qquad \qquad \square$$

Obviously, the last relation can be equivalently expressed in the form

$$Q(v) = \sum_{i=1}^n (\Delta_i/\Delta_{i-1})(v_i)^2, \ \text{ with } \Delta_0 = 1.$$

This last expression has been obtained by Czaki (1970) in a quick way: we assume, as before, that $\Delta_r \neq 0$, for $r = 1, ..., n$, and denote, as before, by $C_{(r)}^{ij}$ the cofactor pertaining to an arbitrary element a_{ij} of the North-West principal submatrix $A_{(r)}$ of order r. Now, let us introduce a new vector $v \in \mathbb{R}^n$, given by the nonsingular transformation

$$x = Tv \text{ or } x^\top = v^\top T^\top,$$

where T is not necessarily orthogonal. The transformation matrix suitable to obtain the desired result is:

$$T = \begin{bmatrix} C_{(1)}^{11}/C_{(1)}^{11} & C_{(2)}^{21}/C_{(2)}^{22} & \cdots & C_{(n)}^{n1}/C_{(n)}^{nn} \\ 0 & C_{(2)}^{22}/C_{(2)}^{22} & \cdots & C_{(n)}^{n2}/C_{(n)}^{nn} \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & C_{(n)}^{nn}/C_{(n)}^{nn} \end{bmatrix}$$

whereas the transposed matrix is, with the symmetry conditions taken into consideration,

$$T^\top = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ C_{(2)}^{12}/C_{(2)}^{22} & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ C_{(n)}^{1n}/C_{(n)}^{nn} & C_{(n)}^{2n} & \cdots & 1 \end{bmatrix},$$

where $C_{(2)}^{12} = C_{(2)}^{21} = -a_{21} = -a_{12}$, $C_{(2)}^{22} = a_{11}$, etc.

After posing $\Delta_0 = 1$ by definition, the original quadratic form $Q(x) = x^\top A x$ can be converted into the following diagonal form:

$$Q(Tv) = v^\top T^\top A T v = v^\top Y v,$$

where

$$Y = T^\top A T = diag\left[\Delta_1/\Delta_0, \Delta_2/\Delta_1, ..., \Delta_{n-1}/\Delta_{n-2}, \Delta_n/\Delta_{n-1}\right],$$

because $C_{(r)}^{rr} = \Delta_{r-1}$ for $r = 1, 2, ..., n$. Thus, the diagonal form can be expressed as

$$Q(v) = \Delta_1(v_1)^2 + \frac{\Delta_2}{\Delta_1}(v_2)^2 + ... + \frac{\Delta_{n-1}}{\Delta_{n-2}}(v_{n-1})^2 + \frac{\Delta_n}{\Delta_{n-1}}(v_n)^2. \qquad \square$$

D) Proofs which use the "Inclusion Principle" for eigenvalues.

Some authors (e. g. Franklin (1968), Horn and Johnson (1985), Johnson (1970)) make speed and short the "second step" of the sufficiency proof sub **A)** by means of the so-called *Inclusion Principle* for eigenvalues (of a symmetric matrix), known also as *Cauchy's Interlace Theorem,* which is in turn a special case of the more general *Courant-Fischer Theorem.* Rather short proofs of Cauchy's Interlace Theorem are given by Hwang (2004) and by Warrington (2005).

Theorem 5   (Inclusion Principle for eigenvalues or Cauchy's Interlace Theorem).

Let $A = \begin{bmatrix} a_{ij} \end{bmatrix}$ be a symmetric matrix of order n, with $n > 1$. Let $A_{(n-1)}$ be the North-West principal submatrix of A of order $(n-1)$. Let $\alpha_1 \geqq \alpha_2 \geqq ... \geqq \alpha_n$ be the eigenvalues of A and let $\beta_1 \geqq \beta_2 \geqq ... \geqq \beta_{n-1}$ be the eigenvalues of $A_{(n-1)}$. Then

$$\alpha_1 \geqq \beta_1 \geqq \alpha_2 \geqq \beta_2 \geqq ... \geqq \alpha_{n-1} \geqq \beta_{n-1} \geqq \alpha_n.$$

With the help of Theorem 5, the sufficient part of the proof of the Sylvester criterion by induction can be simply performed as follows. Suppose that $\Delta_i \equiv \det(A_{(i)}) > 0$ for all $i \leqq n$. Then $A_{(1)} = a_{11}$ is obviously positive definite. Supposing that $A_{(k)}$ is positive definite for some $k < n$, we have to prove that $A_{(k+1)}$ is positive definite. If $\alpha_1 \geqq ... \geqq \alpha_{k+1}$ are the eigenvalues of $A_{(k+1)}$ and if $\beta_1 \geqq ... \geqq \beta_k$ are the eigenvalues of $A_{(k)}$, the Cauchy Interlace Theorem tell us that

$$\alpha_1 \geqq \beta_1 \geqq \alpha_2 \geqq \beta_2 \geqq ... \geqq \alpha_k \geqq \beta_k \geqq \alpha_{k+1}.$$

But all β_i are positive, since $A_{(k)}$ is positive definite, therefore all α_i, $i = 1, ..., k$, are positive. But being

$$\alpha_1 \alpha_2 ... \alpha_k \alpha_{k+1} = \Delta_{k+1} > 0,$$

we conclude that also $\alpha_{k+1} > 0$ and therefore that also $A_{(k+1)}$ is positive definite. $\qquad \square$

E)　"Cyclical Proofs".

Some authors (e. g. Meyer (2000), Nicholson (2003), Strang (2006), Schlesinger (2011)) present "cyclical proofs" of various (equivalent) conditions for the positive definiteness of a symmetric matrix, so obtaining the Sylvester criterion as one of these conditions. Usually, these proofs benefit from the "pivots theory", and especially from the *Cholesky decomposition or Cholesky factorization*: we recall that a symmetric matrix A is positive definite (Theorem 3) if and only if $A = Q^\top Q$ for some nonsingular matrix Q. This matrix Q is not necessarily unique; however, there exists one and only one *upper-triangular* matrix U with positive diagonal elements (and therefore nonsingular) such that

$$A = U^\top U.$$

This is precisely the Cholesky factorization of A.

As a consequence of the Cholesky factorization it is possible to prove that if and only if A is positive definite, it admits a unique factorization of the form

$$A = LDL^\top$$

where L is a lower-triangular matrix, with diagonal elements all equal to one and D is a diagonal matrix, with positive elements on its diagonal. Note that if we take $U = \sqrt{D}L^\top$, where

$$\sqrt{D} = diag(\sqrt{d_1}, ..., \sqrt{d_n}),$$

then U is an upper-triangular matrix with positive elements on its principal diagonal and it results

$$A = U^\top U,$$

i.e. the Cholesky factorization. The matrix $U = \sqrt{D}L^\top$ is also called the *Cholesky factor* of A.

As for what concerns "pivot theory" it can be proved (Meyer (2000), Strang (2006)) that if all leading principal minors of the symmetric matrix A are positive, then the k-th pivot is given by

$$u_{kk} = \begin{cases} \Delta_1 = a_{11} \ \text{for } k = 1 \\ \Delta_k/\Delta_{k-1} \ \text{for } k = 2, 3, ..., n. \end{cases} \tag{3}$$

This result clarifies the relationships between the pivots theory and the Sylvester criterion. Moreover, it can be seen (see, e. g., Meyer (2000), Strang (2006)), that in the above factorization $A = LDL^\top$, the elements d_k of D are just the pivots of A. Without entering into details, which can be found in the above quoted books, a possible "cyclical" theorem could be the following one.

Theorem 6.　For real symmetric matrices A, of order n, the following statements are equivalent.

1)　$Q(x) = x^\top A x > 0$, $\forall x \in \mathbb{R}^n$, $x \neq 0$.
2)　All eigenvalues of A are positive.
3)　$A = B^\top B$ for some nonsingular matrix B.
4)　There exists a unique upper-triangular matrix U with positive diagonals such that
　　$A = U^\top U$ ("Cholesky factorization of A").
5)　A admits a factorization of the form $A = LDL^\top = U^\top U$, where $U = D^{\frac{1}{2}}L^\top$ and where
　　the elements d_k of the diagonal matrix D are all positive.
6)　All pivots of A (without row exchange) are positive.
7)　The leading principal minors of A are positive.
8)　All principal minors of A are positive.

Remark 2.　For what concerns the main subject of the present Section, i. e. the Sylvester criterion, the same can be obtained without going along all the steps of the previous theorem. For example, first it is convenient to prove that A has all positive eigenvalues if and only if $A = B^\top B$ for some nonsingular matrix B. Then, the expression $A = B^\top B$ is equivalent to the fact that A possesses a factorizarion of the form $A = U^\top U$, where U is an upper triangular matrix with positive diagonal entries and where all pivots of A are positive and given by expression (3).

F)　Proofs based on properties of linear spaces.

Some proofs of the Sylvester criterion (see, e. g., Gilbert (1991), Thrall and Tornheim (1957)) are based on "geometric" properties of linear spaces; these proofs are in general quite concise, however not too suitable for a course of Linear Algebra for undergraduates. The elegant proof of Gilbert (1991) can be schematized as follows. As already noted before,

the necessity part of the proof of the Sylvester criterion shows no particular difficulty. For the sufficient part we need the following two lemmas.

Lemma 2.   Let $v^1, ..., v^n$ be a basis of a finite-dimensional vector space V. Suppose W is a k-dimensional subspace of V. If $m < k$, then there exists a nonzero vector in W which belongs to the span of $v^{m+1}, ..., v^n$.

Proof.   Let $w^1, ..., w^k$ be a basis for W. If the statement of the lemma is false, then $w^1, ..., w^k, v^{m+1}, ..., v^n$ is a basis of V, which contradicts that dim $V = n$.   □

Lemma 3.   Let A be a symmetric (real) matrix of order n. Suppose W is a k-dimensional subspace of $\mathbb{R}^n$ such that

$$w^\top A w > 0, \ \forall w \in W, \ w \neq 0.$$

Then A has at least k (counting multiplicity) positive eigenvalues.

Proof.   By the spectral theorem, there is an orthonormal basis $v^1, ..., v^n$ for $\mathbb{R}^n$ consisting of eigenvectors of A with corresponding eigenvalues $\lambda_1, ..., \lambda_n$. Suppose that the first m eigenvalues are positive and the remaining not. If $m < k$, then Lemma 2 implies there exists a nonzero vector w in W which may be written

$$w = a_{m+1} v^{m+1} + ... + a_n v^n.$$

Applying our hypothesis that the quadratic form defined by A is positive on W, we obtain

$$w^\top A w = A w \cdot w = (a_{m+1} \lambda_{m+1} v^{m+1} + ... + a_n \lambda_n v^n) \cdot (a_{m+1} v^{m+1} + ... + a_n v^n) =$$

$$= \lambda_{m+1} a_{m+1}^2 + ... + \lambda_n a_n^2 > 0.$$

However, we assumed that $\lambda_{m+1}, ..., \lambda_n$ are all non positive, so that

$$\sum_{i=m+1}^{n} \lambda_i a_i^2 \leqq 0,$$

which is a contradiction. Thus $m \geqq k$.   □

Now, it is possible to prove the sufficiency of the Sylvester criterion by induction on n. For $n = 1$, there is nothing to prove, as we already know. Suppose that the sufficiency of Sylvester's criterion is true for the $(n-1) \times (n-1)$ leading principal submatrix of A, i. e. for $A_{(n-1)}$. Let W denote the subspace of $\mathbb{R}^n$ consisting of all vectors with zero n-th coordinate. Let P denote the projection of $\mathbb{R}^n$ onto the subspace spanned by the first $(n-1)$ coordinates with respect to the standard basis. Then

$$w^\top A w = (w^\top P) A (P^\top w) = w^\top A_{(n-1)} w > 0, \ \forall w \in W, \ w \neq 0$$

by our induction hypothesis. By the preceding lemma, A has at least $(n-1)$ positive eigenvalues (counting multiplicity). Therefore, the n-th eigenvalue is also positive, since otherwise, the n-th leading principal minor, i. e. $\det(A)$, would be non positive, which is a contradiction.   □

3. The Sylvester Criterion for Quadratic Forms Subject to Linear Constraints

In many applications of real quadratic forms, especially in optimization problems with equality constraints, we are interested in the sign of the quadratic form (1), where $x \in S$, S being the set of all non zero solutions to the homogeneous system of linear equations

$$Bx = 0,$$

where B is a (real) (m, n) matrix, with $m < n$.

The related extension of the Sylvester criterion for the above problem has been treated by many authors; see, e. g., Afriat (1951), Bellman (1970), Chabrillac and Crouzeix (1984), Debreu (1952), Farebrother (1977), Lancaster (1968), Mann (1943), Murata (1977), Samuelson (1947). Perhaps the first paper which extends the Sylvester criterion to the case under examination is the paper of Mann (1943), where the related proof is however quite long and with many algebraic transformations (there is also a misprint in the second one of Mann's conditions). The treatment of Debreu (1952) is sound and rather stringent, however this author makes use of tools of Mathematical Analysis, whereas the problem into consideration is a typical algebraic problem. Chabrillac and Crouzeix (1984) and Farebrother (1977) employ the Schur's complement, the inertia law and a classical result of Finsler. The treatments of Lancaster (1968) and Samuelson (1947) are interesting, but not complete and rather unsatisfactory. The papers of Black and Morimoto (1968) and Väliaho (1982)

are concise and in the same spirit of our proof. The paper of Afriat (1951) is rather intricate; the proof of Truchon (1987) expands the concise proof of Debreu (1952) and is therefore quite long. McFadden (1978) presents some equivalent determinantal conditions for a symmetric matrix to be positive definite subject to homogeneous linear constraints. There are some obscure points: e. g. it is stated that the matrix

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

has a positive principal minor of each order. Moreover, the conditions of Mc Fadden are in general less operative than the Sylvester conditions. Two other interesting papers are the one of Cherubino (1957), in Italian, and the one of Shostak (1954), in Russian.

We try to give an algebraic proof of the said generalization of the Sylvester criterion, proof which uses only classical results of Linear Algebra. We base our treatment on the paper of Giorgi (2003). Let us consider the quadratic form (1):

$$Q(x) = x^\top A x$$

(*A* symmetric matrix of order *n*), subject to the linear constraints

$$Bx = 0, \tag{4}$$

where *B* is of order (m, n), with $m < n$ and

$$rank(B) = m.$$

Theorem 7.   Without loss of generality, let us suppose that the *first m* columns of *B* are linearly independent. Then:

(*i*) $Q(x) = x^\top A x > 0$ for all $x \neq 0$ such that $Bx = 0$ if and only if the leading principal minors of order $2m + p$, $p = 1, 2, ..., n - m$, of the matrix

$$H = \begin{bmatrix} 0 & B \\ B^\top & A \end{bmatrix}$$

have the sign of $(-1)^m$. Denoting by $\tilde{\Delta}_k$ the *k*-th leading principal minor of *H*, we must therefore have

$$(-1)^m \tilde{\Delta}_k > 0, \ k = 2m + 1, ..., m + n.$$

(*ii*) $Q(x) = x^\top A x < 0$ for all $x \neq 0$ such that $Bx = 0$ if and only if the leading principal minors of *H*, of order $2m + p$, $p = 1, 2, ..., n - m$, have the sign of $(-1)^{m+p}$.

Remark 3.   Also in the previous theorem statement (*ii*) derives from statement (*i*) by observing that *A* is negative definite under constraints if and only if $-A$ is positive definite under the same constraints.

Remark 4.   In some papers and books the following bordered matrix is considered:

$$\bar{H} = \begin{bmatrix} A & B^\top \\ B & 0 \end{bmatrix}.$$

In this case obviously the previous conditions (*i*) and (*ii*) have to be suitably modified, by making reference to the leading principal minors of $\bar{H}$.

Remark 5.   In case of only one constraint, of the type $bx = 0$, with $b \in \mathbb{R}^n$ and with $b_1 \neq 0$, the previous conditions (*i*) and (*ii*) of Theorem 7 become, respectively,

a) $Q(x)$ is positive definite on the set of non trivial solutions of $bx = 0$ if and only if

$$\tilde{\Delta}_3 = \begin{vmatrix} 0 & b_1 & b_2 \\ b_1 & a_{11} & a_{12} \\ b_2 & a_{21} & a_{22} \end{vmatrix} < 0, ..., |H| < 0.$$

b) $Q(x)$ is negative definite on the set of non trivial solutions of $bx = 0$ if and only if

$$\tilde{\Delta}_3 = \begin{vmatrix} 0 & b_1 & b_2 \\ b_1 & a_{11} & a_{12} \\ b_2 & a_{21} & a_{22} \end{vmatrix} > 0, \quad \tilde{\Delta}_4 = \begin{vmatrix} 0 & b_1 & b_2 & b_3 \\ b_1 & a_{11} & a_{12} & a_{13} \\ b_2 & a_{21} & a_{22} & a_{23} \\ b_3 & a_{31} & a_{32} & a_{33} \end{vmatrix} < 0, \ \text{etc.}$$

Remark 6.   We have to note that the assumption that the rank of B ($rank(B) = m$) is given by the first m columns of B is essential to have necessary and sufficient conditions. In absence of this assumption, conditions (*i*) and (*ii*) of Theorem 7 are only sufficient conditions and they imply $rank(B) = m$. This is the case of the usual sufficient second-order optimality conditions imposed on the Lagrangian function of optimization problems with equality constraints. Consider, e. g., the following simple example.

It is obvious that the quadratic form

$$Q(x) = (x_1)^2 + (x_2)^2 - (x_3)^2$$

is positive definite on the constraint $x_3 = 0$ (hence here vector b is $b = [0, 0, 1]$). Yet we have

$$\tilde{\Delta}_3 = \begin{vmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 0.$$

Proof of Theorem 7.   On the grounds of Remark 3, we prove only statement (*i*). Let us partition matrix B into two submatrices B_1 and B_2, where B_1 contains the first m columns of B, so that $\det(B_1) \neq 0$, and B_2 contains the last $(n - m)$ columns of B :

$$B = [B_1, B_2].$$

Similarly, we separate vector $x \in \mathbb{R}^n$ into two blocks:

$$\begin{aligned} x^1 &= [x_1, x_2, ..., x_m]^\top ; \\ x^2 &= [x_{m+1}, x_{m+2}, ..., x_n]^\top . \end{aligned}$$

Finally, we partition matrix A into four submatrices, as follows:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

where $A_{21} = A_{12}^\top$, A_{11} is symmetric of order m, A_{12} is of order $(m, (n - m))$ and A_{22} is symmetric of order $(n - m)$.

Now we compute $Q(x)$, taking into account that $x = \begin{bmatrix} x^1, x^2 \end{bmatrix}^\top$:

$$Q(x) = x^\top A x = \begin{bmatrix} x^1, x^2 \end{bmatrix}^\top \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} =$$

$$= (x^1)^\top A_{11} x^1 + (x^1)^\top A_{12} x^2 + (x^2)^\top A_{12}^\top x^1 + (x^2)^\top A_{22} x^2.$$

The m variables $x_1, x_2, ..., x_m$ can be eliminated, as x must satisfy relation (4). Taking into account that $x^1 = -B_1^{-1} B_2 x^2$, we find

$$Q(x^2) = Q_2(x^2) = (x^2)^\top A^* x^2,$$

where A^* is the symmetric matrix given by

$$A^* = B_2^\top (B_1^\top)^{-1} A_{11} B_1^{-1} B_2 - B_2^\top (B_1^\top)^{-1} A_{12} - A_{12}^\top B_1^{-1} B_2 + A_{22}.$$

Putting $J = -B_1^{-1} B_2$ we can also compact the above expression:

$$A^* = J^\top A_{11} J + J^\top A_{12} + A_{12}^\top J + A_{22}.$$

In other words, the sign of $Q(x)$ under the constraints $Bx = 0$ coincides with the sign of $Q_2(x)$, without constraints. Then, it is possible to establish a congruence relation between two symmetric matrices, which allows to compare the sign of the North-West principal minors of the square matrix A^* (of order $n - m$) with the North-West principal minors of the square matrix H.

More precisely, we form the following matrices C, M and N, defined as follows:

$$C = \begin{bmatrix} I_m & 0 & 0 \\ 0 & B_1 & B_2 \\ 0 & 0 & I_{n-m} \end{bmatrix},$$

where I_k is the identity matrix of order k;

$$M = (B_1^\top)^{-1} A_{11} B_1^{-1};$$

$$N = (B_1^\top)^{-1} A_{12} - (B_1^\top)^{-1} A_{11} B_1^{-1} B_2.$$

It is quite easy to verify that

$$C^\top P C = H,$$

where P is the symmetric matrix

$$P = \begin{bmatrix} 0 & I_m & 0 \\ I_m & M & N \\ 0 & N^\top & A^* \end{bmatrix}.$$

By performing, by blocks, the product $C^\top P C$, we realize that for every $p = 1, 2, ..., n - m$, the North-West principal minor $\tilde{\Delta}_{2m+p}$ of order $2m + p$ of H is linked to the North-West principal minor Δ_p^* of order p of A^* by the formula

$$\tilde{\Delta}_{2m+p} = (-1)^m |B_1|^2 \Delta_p^*.$$

We obtain therefore that the signs of Δ_p^* and $(-1)^m \tilde{\Delta}_{2m+p}$ coincide, for $p = 1, 2, ..., n - m$. □

Remark 7.　　The case where the constraints of the quadratic form are given by a non-homogeneous system of the type

$$Bx = c, \tag{5}$$

is studied by Shostak (1954). Also Heal, Hughes and Tarling (1974) consider a constraint system of the type (5), but then they treat (in an unsatisfactory way) only the case $c = 0$. For the reader's convenience we report the results of Shostak. Again we consider the quadratic form (1) subject to (5), where B is of order (m, n), $m < n$, $rank(B) = m$ and where the determinant of the square submatrix B_1, formed by the first m columns of B, is different from zero. Let us transform the variables in (1) and in (5) by means of the following position:

$$x_j = \frac{y_j}{y_{n+1}}, \ \ j = 1, ..., n, \ (y_{n+1} \neq 0).$$

Then (1) becomes

$$Q'(y_1, ..., y_{n+1}) = \frac{1}{y_{n+1}^2} \sum_{i,j=1}^{n} a_{ij} y_i y_j \tag{6}$$

and (5) becomes

$$\sum_{j=1}^{n} b_{ij} y_j - c_i y_{n+1}, \ i = 1, ..., m. \tag{7}$$

The problem is then of characterizing the sign of (6) under the constraints (7). The positive or negative definiteness of the said constrained quadratic form depends from the sign of the North-West principal minors, of order $2m + 1, ..., m + n + 1$, of the bordered matrix

$$H_c = \begin{bmatrix} 0 & B & c \\ B^\top & A & 0 \\ c^\top & 0 & 0 \end{bmatrix}.$$

References

Afriat, S. N. (1951). The quadratic form positive definite on a linear manifold, *Proc. Cambridge Phil. Soc.,* **47,** 1-6. https://doi.org/10.1017/S0305004100026293

Aleskerov, F., Ersel, M., & Piontkovski, D. (2011). Linear Algebra for Economists, Springer, Berlin. https://doi.org/10.1007/978-3-642-20570-5

R. BELLMAN (1970). Introduction to Matrix Analysis, McGraw-Hill Book Co., New York.

Black, J., & Morimoto, Y. (1968). A note on quadratic forms positive definite under linear constraints, *Economica,* **35,** 205-206. https://doi.org/10.2307/2552137

Bomze, T. M. (1995). Checking positive-definiteness by three statements, *Int. J. Math. Educ. Sci. Technol.,* **26,** 289-294.

Chabrillac, Y., & Crouzeix, J.-P. (1984). Definiteness and semidefiniteness of quadratic forms revisited, *Linear Algebra and Its Appl.,* **63,** 283-292. https://doi.org/10.1016/0024-3795(84)90150-2

Cherubino, S. (1957). Forme quadratiche (o hermitiane) subordinanti forme definite o semidefinite, *L'Industria,* **N. 3,** 500-504.

Czaki, F. (1970). A concise proof of Sylvester's theorem, *Periodica Polytechnica Electrical Engineering,* **14,** 105-112.

G. Debreu, (1952). Definite and semidefinite quadratic forms, *Econometrica,* **20,** 295-300. https://doi.org/10.2307/1907852

Farebrother, R. W. (1977). Necessary and sufficient conditions for a quadratic form to be positive whenever a set of homogeneous linear constraints is satisfied, *Linear Algebra and Its Appl.,* **16,** 39-42. https://doi.org/10.1016/0024-3795(77)90017-9

Franklin, J. N. (1968). Matrix Theory, Prentice-Hall, Englewood Cliffs, N. J.

Gantmacher, F. R. (1959). The Theory of Matrices, Vol. 1, Vol. 2, Chelsea Publishing Co., New York.

Gilbert, G. T. (1991). Positive definite matrices and Sylvester's criterion, *American Math. Monthly,* **98,** 44-46. https://doi.org/10.2307/2324036

Giorgi, G. (2003). Note sulle forme quadratiche libere e condizionate e sul riconoscimento del segno tramite minori. In: G. Crespi and others (Eds.). Recent Advances in Optimization, DATANOVA, Milan, 71-86.

Heal, G., Hughes, G., & Tarling, R. (1974). Linear Algebra and Linear Economics, The Macmillan Press Ltd, London.

Hestenes, M. R. (1975). Optimization Theory: The Finite-Dimensional Case, J. Wiley, New York.

Hohn, F. E. (1973). Elementary Matrix Algebra, The Macmillan Company, New York.

Horn, R. A., & Johnson, C. R. (1985). Matrix Analysis, Cambridge Univ. Press, Cambridge. https://doi.org/10.1017/CBO9780511810817

Hwang, S.-G. (2004). Cauchy's interlace theorem for eigenvalues of Hermitian matrices, *American Math. Monthly,* **111,** 157-159. https://doi.org/10.2307/4145217

Johnson, C. R. (1970). Positive definite matrices, *American Math. Monthly,* **77,** 259-264. https://doi.org/10.2307/2317709

Kemp, M. C., & Kimura, Y. (1978). Introduction to Mathematical Economics, Springer-Verlag, New York. https://doi.org/10.1007/978-1-4612-6278-7

Lancaster, K. (1968). Mathematical Economics, Macmillan, London.

Mann, H. B. (1943). Quadratic forms with linear constraints, *American Math. Monthly,* **50,** 430-433. https://doi.org/10.2307/2303666

Mcfadden, D. (1978). Definite quadratic forms subject to constraints. In: M. Fuss & D. McFadden, Production Economics - A Dual Approach to Theory and Applications, Vol. I, North Holland, Amsterdam.

Meyer, C. (2000). Matrix Analysis and Applied Linear Algebra, SIAM, Fhiladelphia.

Mirsky, L. (1955). An Introduction to Linear Algebra, Clarendon Press, Oxford.

Murata, Y. (1977). Mathematics for Stability and Optimization of Economic Systems, Academic Press, New York.

Nicholson, W. K. (2003). Elementary Linear Algebra, McGraw-Hill Education, New York.

Samuels, S. M. (1966). A simplified proof of a sufficient condition for a positive definite quadratic form, *American Math. Monthly,* **73,** 297-298. https://doi.org/10.2307/2315352

Samuelson, P. A. (1947). Foundations of Economic Analysis, Harvard Univ. Press, Cambridge, Mass.

Schlesinger, E. (2011). Algebra Lineare e Geometria, Zanichelli, Bologna.

Shostak, R. Y. A. (1954). On a criterion of conditional definiteness of a quadratic form of *n* variables subject to linear relations and on a sufficient condition for a conditional extremum of a function of *n* variables, (in Russian), *Uspekhi Mat. Nauk,* **9,** 199-206.

Simon, C., & Blume, L. (1994). Mathematics for Economists, W. W. Norton & Co., New York.

Strang, G. (2006). Linear Algebra and Its Applications, Brookes/Cole Cengage Learning.

Thrall, R. M., & Tornheim, L. (1957). Vector Spaces and Matrices, J. Wiley & Sons, New York. https://doi.org/10.1063/1.3060170

Truchon, M. (1987). Théorie de l'Optimisation Statique et Différentiable, Gaëtan Morin Editeur, Chicoutimi, Québec, Canada.

H. VäLiaho (1982). On the definity of quadratic forms subject to linear constraints, *J. Optim. Theory Appl.,* **38,** 143-145. https://doi.org/10.1007/BF00934328

Warrington, G. S. (2005). A very short proof of Cauchy's interlace theorem, *American Math. Monthly,* **112,** 118.

Copyrights