Addendum on Quadratic Forms

Michael H. Mertens

December 10, 2015

1 Introduction

In these short notes, I intend to make some of the material on quadratic forms and symmetric matrices that was covered in the lecture but which is not contained in the textbook. The notation will be as in the lecture, the numbering of theorems and definitions is not (for technical reasons). Usually, the letter Q will denote a quadratic form, its associated matrix (which is always symmetric) is usually denoted by the letter A. The set of all symmetric $n \times n$-matrices is denoted by $\mathbb{R}^{n\times n}_{sym}$.

Should there be any typos or mathematical errors in this manuscript, I'd be glad to hear about them via email (`michael.mertens@emory.edu`) and correct them. Please note that the numbering of theorems and definitions is unfortunately not consistent with the lecture.

Atlanta, November 2015, Michael H. Mertens

2 The signature and Sylvester's Inertia Theorem

Recall the following definition.

Definition 2.1. *Let Q_1 and Q_2 be quadratic forms in n variables with $Q_j(x) = x^{tr}A_jx$ with $A_j \in \mathbb{R}^{n\times n}_{sym}$ for $j = 1, 2$. We call the two quadratic form Q_1 and Q_2 (resp. the symmetric matrices A_1 and A_2)* equivalent*, if there is an invertible matrix $g \in \mathbb{R}^{n\times n}$ such that*

$$A_1 = g^{tr}A_2g.$$

Remark 2.2. *Equivalence of quadratic forms corresponds to a change of variables. If we have (in the notation of Definition 2.1) $A_1 = g^{tr}A_2g$, then we have for all $x \in \mathbb{R}^n$ that*

$$Q_1(x) = x^{tr}A_1x = x^{tr}g^{tr}A_2gx = (gx)^{tr}A_2(gx) = Q_2(gx).$$

Hence two quadratic forms take the same values if and only if they are equivalent.

There is a convenient way to determine whether two quadratic forms are equivalent or not. This is the content of the following theorem, which is widely known as Sylvester's Inertia Theorem.

Theorem/Definition 2.3 (Sylvester's Inertia Theorem)**.** *Let Q be a quadratic form in n variables with associated matrix $A \in \mathbb{R}^{n\times n}_{sym}$. Then the following are true.*

(i) Q is equivalent to a quadratic form whose associated matrix is given by

$$\mathrm{diag}(1, ..., 1, -1, ..., -1, 0, ..., 0) \in \mathbb{R}^{n\times n}_{sym}.$$

Denoting the number of 1's in the above matrix by a_+, the number of -1's by a_- and the number of 0's on the diagonal by a_0, we call the triple $\mathrm{sig}(A) := (a_+, a_-, a_0)$ *the* signature *of A (respectively Q).*

(ii) Two quadratic forms are equivalent if and only if they have the same signature.

Proof. By the Principal Axes Theorem, Q is (orthogonally) equivalent to a form whose matrix is diagonal, i.e., there is an orthogonal (in particular invertible) matrix $g \in \mathbb{R}^{n\times n}$ such that

$$A = g^{tr} D g,$$

where $D = \operatorname{diag}(\lambda_1, ..., \lambda_n)$ and $\lambda_1, ..., \lambda_n$ are the eigenvalues of A, with repetitions according their algebraic multiplicities. Without loss of generality we can assume that the eigenvalues are sorted such that the first a_+ of them are positive, the following a_- are negative, and the remaining a_0 eigenvalues are 0. Then we define the matrix

$$C = \operatorname{diag}(\sqrt{|\lambda_1|}, ..., \sqrt{|\lambda_{a_+ + a_-}|}, 1, ..., 1).$$

This matrix is clearly invertible and we have

$$D = C \operatorname{diag}(1, ..., 1, -1, ..., -1, 0, ..., 0) C,$$

hence

$$A = (g^{tr} C) \operatorname{diag}(1, ..., 1, -1, ..., -1, 0, ..., 0)(Cg) = (Cg)^{tr} \operatorname{diag}(1, ..., 1, -1, ..., -1, 0, ..., 0)(Cg),$$

so that the first claim follows.

This follows directly from the fact that equivalence is *transitive*, i.e., if Q_1 and Q_2 are equivalent and Q_2 and Q_3 are equivalent, thenQ_1 and Q_3 are equivalent as well. □

Example 2.4. *Compute the signature of the quadratic form*

$$Q(x) = -x_1^2 - 6x_1x_2 - 2x_1x_3 + 14x_2^2 + 16x_2x_3 + 11x_3^2.$$

The matrix of Q is given by

$$A = \begin{pmatrix} -1 & -3 & -1 \\ -3 & 14 & 8 \\ -1 & 8 & 11 \end{pmatrix}.$$

From the proof of the Inertia Theorem, we see that one way to find the signature of A is to compute the eigenvalues of A and determine their signs. This is however not the easiest thing to do since the minimal polynomial *of A has no "nice" zeros (the eigenvalues of A are approximately given by* $-1.61047...,\ 4.57546...,\ 21.035005...$*).*

The following method is usually much easier to apply than the computation of eigenvalues. Recall that performing a row reduction on the matrix A corresponds to multiplying A from the left by an invertible matrix g. It is not hard to see that multiplying A from the right by g^{tr}. Now it is obvious that A and gAg^{tr} are similar. Hence we can use simultaneous row- and column reduction to transform A into diagonal form from which we can easily read off the signature (from the signs of the diagonal entries). It is most important that one always performs the "same" column operation as the previous row operation. Note that this method does in general **NOT** *preserve the eigenvalues of A, but only their signs.*

$$A \to \begin{pmatrix} -1 & -3 & -1 \\ 0 & 23 & 11 \\ 0 & 11 & 12 \end{pmatrix}$$

$$\downarrow$$

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 23 & 11 \\ 0 & 11 & 12 \end{pmatrix} \to \begin{pmatrix} -1 & 0 & 0 \\ 0 & 23 & 11 \\ 0 & 0 & \frac{155}{23} \end{pmatrix}$$

$$\downarrow$$

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 23 & 0 \\ 0 & 0 & \frac{155}{23} \end{pmatrix}$$

From here we see immediately that $\operatorname{sig}(A) = (2, 1, 0)$.

Another way to determine the signature of a symmetric matrix was discovered by Adolf Hurwitz. We need one further definition.

Definition 2.5. *Let $B \in \mathbb{R}^{n\times n}$. Then we call*

$$\delta_j(B) := \det \begin{pmatrix} b_{11} & \dots & b_{1j} \\ \vdots & \ddots & \vdots \\ b_{j1} & \dots & b_{jj} \end{pmatrix}$$

the j^{th} principal minor *of B. We also define the 0^{th} principal minor to be* $\delta_0(B) := 1$.

Theorem 2.6 (Hurwitz criterion)**.** *Let Q be a quadratic form with matrix $A \in \mathbb{R}^{n\times n}_{sym}$ whose principal minors are all non-zero. Then we have* $\operatorname{sig}(Q) = (n-q, q, 0)$*, where q is the number of sign changes in the sequence* $(\delta_0(A), ..., \delta_n(A))$ *of the principal minors of A.*

Proof. Suppose that $\operatorname{sig}(A) = (a_+, a_-, a_0)$. By the assumption that all principal minors of A, so in particular $\delta_n(A) = \det A$, are non-zero, we see that our form is non-degenerate, i.e., $a_0 = 0$. Furthermore, it follows from the Inertia Theorem 2.3 that $\det A = (-1)^{a_-}|\det A|$.

With these preliminary remarks, we can now carry out the proof using mathematical induction. The claim is clear for $n = 1$: Then we have $\delta_0 = 1$ and $\delta_1 = a_{11} \neq 0$. If $a_{11} > 0$, we have no sign changes and the signature is $(1, 0, 0)$, if $a_{11} < 0$ then there is one sign change and the signature is $(0, 1, 0)$.

Now assume that the claim is true for some natural number n. We need to show that it is then also true for $n + 1$. So let $A \in \mathbb{R}^{(n+1)\times(n+1)}_{sym}$ with signature $(a_+, a_-, 0)$, whose principal minors are all non-zero. Then we have $a_+ = n + 1 - a_-$. Now define the matrix $A' \in \mathbb{R}^{n\times n}_{sym}$ by deleting the last row and column of A, this matrix will have signature $(n - a'_-, a'_-, 0)$, where we know, bu the induction hypothesis, that a'_- is the number of sign changes in the sequence $(\delta_0(A'), ..., \delta_n(A')) = (\delta_0(A), ..., \delta_n(A))$. Then we have $n - a'_- \leq n - a_-$ and $a'_- \leq a_-$, which implies

$$a'_- \leq a_- \leq a'_- + 1 \tag{2.1}$$

Now we distinguish two cases.

(*i*) $\delta_{n+1}(A)$ and $\delta_n(A)$ have the same sign. Then we have by the preliminary remark that $(-1)^{a'_-} = (-1)^{a_-}$, which implies with (2.1) that $a'_- = a_-$, which is the number of sign changes in $(\delta_0(A), ..., \delta_{n+1}(A))$.

(*ii*) $\delta_{n+1}(A)$ and $\delta_n(A)$ have opposite signs. Then $(-1)^{a'_-} = -(-1)^{a_-}$, which implies together with (2.1) that $a_- = a'_- + 1$, which is again the number of sign changes in $(\delta_0(A), ..., \delta_{n+1}(A))$.

This completes the proof. $\square$

Example 2.7. *Compute the signature of the matrix*

$$A = \begin{pmatrix} 2 & 1 & -3 \\ 1 & 3 & 0 \\ -3 & 0 & -2 \end{pmatrix} \in \mathbb{R}^{3\times 3}_{sym}.$$

We employ the Hurwitz criterion and compute the principal minors,

$$\delta_0(A) = 1,\ \delta_1(A) = 2,\ \delta_2(A) = \det \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} = 5,\ \delta_3(A) = \det A = -37.$$

Since the sequence $(1, 2, 5, -37)$ *contains exactly one sign change, we find by Theorem 2.6 that*

$$\operatorname{sig}(A) = (2, 1, 0).$$

3 Solution sets

Quadratic forms are by definition maps form some $\mathbb{R}^n$ to $\mathbb{R}$. A natural question to ask is whether for a given quadratic form Q and real number c, the equation

$$Q(x) = c$$

has a solution and what the solution set looks like. In this section we want to answer this question at least partly and list all possible cases that arise for quadratic forms in 2 and 3 variables. By the Inertia Theorem 2.3 and the following remark, we can regard quadratic forms up to equivalence (since equivalence is merely a linear change of variables, that won't change the basic geometry of our solution set), hence only by their signatures. Moreover, since for $\operatorname{sig}(Q) = (a_+, a_-, a_0)$ we have $\operatorname{sig}((-1)Q) = (a_-, a_+, a_0)$ (which is easy to see) and

$$Q(x) = c \quad \Leftrightarrow \quad -Q(x) = -c,$$

it is enough to consider only signatures (a_+, a_-, a_0) with $a_+ \geq a_1$.

The following concepts play an important role in deciding the solubility of the equation $Q(x) = c$.

Definition 3.1. *let Q be a quadratic form with associated matrix $A \in \mathbb{R}^{n\times n}_{sym}$. Then we call Q (resp. A)*

(i) positive definite *if* $\operatorname{sig}(A) = (n, 0, 0)$*,*

(ii) negative definite *if* $\operatorname{sig}(A) = (0, n, 0)$*,*

(iii) degenerate *if* $\operatorname{sig}(A) = (a_+, a_-, a_0)$ *with* $a_0 > 0$*,*

(iv) indefinite *if* $\operatorname{sig}(A) = (a_+, a_-, a_0)$ *with* $a_+, a_- > 0$*,*

(v) positive semidefinite *if* $\operatorname{sig}(A) = (a_+, 0, a_0)$,

(vi) negative semidefinite *if* $\operatorname{sig}(A) = (0, a_-, a_0)$.

Remark 3.2. *From the proof of the Inertia Theorem it is clear that* $\operatorname{sig}(A) = (a_+, a_-, a_0)$*, where* a_+ *(resp.* a_-*) equals the number of positive (resp. negative) eigenvalues of* A*, and* a_0 *equals the number of* 0 *eigenvalues. Thus* A *is positive (resp. negative) definite if and only if all its eigenvalues are positive (resp. negative), etc.*

Proposition 3.3. *Let* Q *be a quadratic form in* n *variables. Then the following are all true.*

(i) Q *is positive (resp. negative) definite if and only if* $Q(x) > 0$ *(resp.* $Q(x) < 0$*) for all* $x \in \mathbb{R}^n$*,* $x \neq 0$.

(ii) Q *is positive (resp. negative) semidefinite if and only if* $Q(x) \geq 0$ *(resp.* $Q(x) \leq 0$*) for all* $x \in \mathbb{R}^n$.

Proof. Let $A \in \mathbb{R}^{n\times n}_{sym}$ be the matrix of Q. Then we know by the Spectral Theorem (resp the Principal Axes Theorem) that there is an orthonormal basis $\mathscr{B} = \{b_1, ..., b_n\}$ of $\mathbb{R}^n$ consisting of eigenvectors of A. Assume that b_j is an eigenvector to the eigenvalue λ_j.

Writing $x \in \mathbb{R}^n$ in the basis $\mathscr{B}$, we obtain

$$x = \alpha_1 b_1 + ... + \alpha_n b_n$$

for some real weights $\alpha_1, ..., \alpha_n$. Then we have

$$\begin{aligned} Q(x) = x^{tr}Ax =&(\alpha_1 b_1 + ... + \alpha_n b_n)^{tr}(\alpha_1\lambda_1 b_1 + ... + \alpha_n\lambda_n b_n) \\ \overset{\mathscr{B}\ \text{ONB}}{=}&\alpha_1^2\lambda_1 + ... + \alpha_n^2\lambda_n. \end{aligned}$$

Now Q is positive definite if and only if $\lambda_j > 0$ for all j by Remark 3.2, hence in this case we have $x^{tr}Ax > 0$ for all $x \neq 0$ (since at least one $\alpha_i \neq 0$. If, on the other hand, we have, say $\lambda_j = 0$, then $Q(b_j) = 0$, but $b_j \neq 0$, or if $\lambda_1 > 0$ and $\lambda_2 < 0$, then $Q(b_1) > 0$ and $Q(b_2) < 0$, in particular, the inequality $Q(x) > 0$ for all $x \neq 0$ does not hold.

By the same argument, one also obtains the remaining cases. □

As an immediate consequence of Proposition 3.3 we obtain

Corollary 3.4. *If Q is positive (resp. negative) definite, then the equation $Q(x) = c$ has no solutions for $c < 0$ (resp. $c > 0$) and precisely one solution $x = 0$ for $c = 0$.*

We conclude by an enumeration of the possibilities for the solution sets of the equation $Q(x) = c$ for Q in 2 and 3 variables.

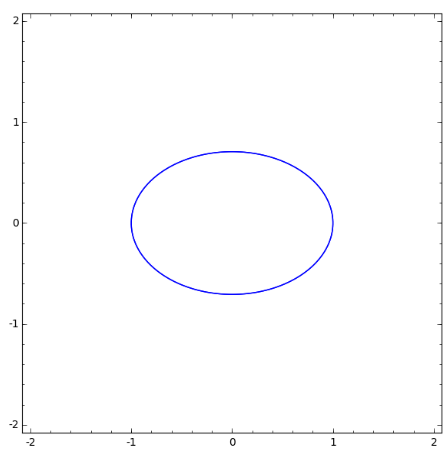
Two variables.

- $\mathrm{sig}(Q) = (2, 0, 0)$. The representative equation is given by

  $$x_1^2 + x_2^2 = c.$$

  The solution set here is

  - an **ellipse** if $c > 0$,

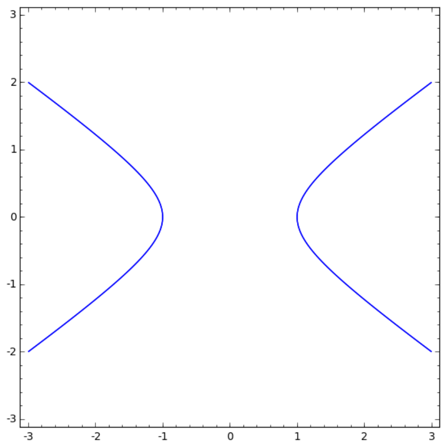
  - a **point** if $c = 0$,
  - empty if $c < 0$.

- $\mathrm{sig}(Q) = (1, 1, 0)$. The representative equation is given by

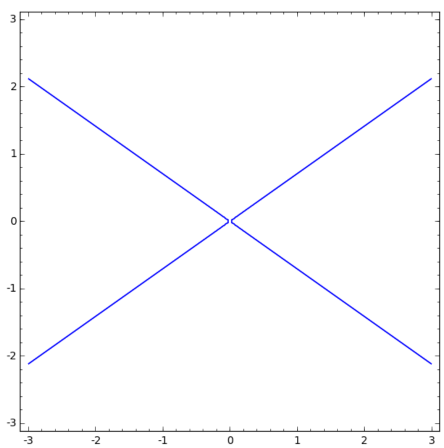
  $$x_1^2 - x_2^2 = c.$$

  The solution set here is

  - a **pair of hyperbolas** if $c \neq 0$,



– a **pair of intersecting lines** if $c = 0$.

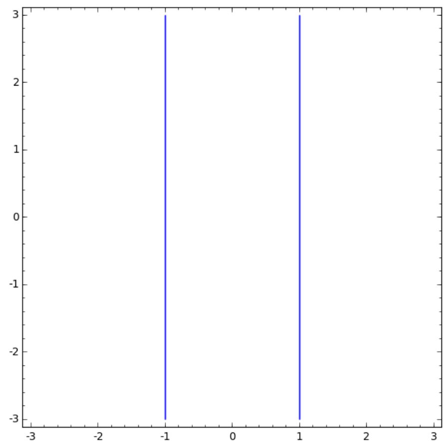


- $\mathrm{sig}(Q) = (1, 0, 1)$. The representative equation is given by

$$x_1^2 = c.$$

  The solution set here is

  – an **pair of parallel lines** if $c > 0$,



- a **line** if $c = 0$,
- empty if $c < 0$.

- $\text{sig}(Q) = (0, 0, 2)$. The representative equation is given by

$$0 = c.$$

The solution set here is

- the full **plane** if $c = 0$,
- empty if $c \neq 0$.

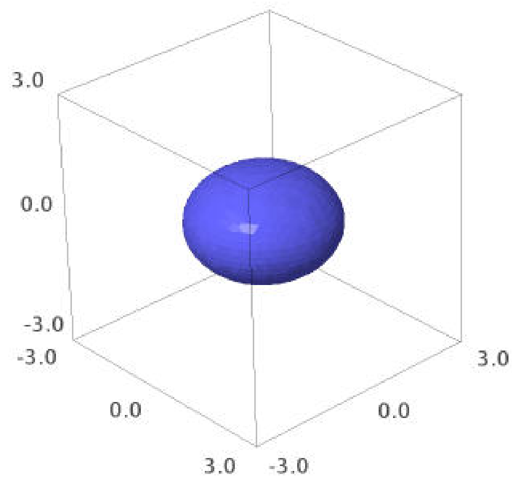
Three variables

- $\text{sig}(Q) = (3, 0, 0)$. The representative equation is given by

$$x_1^2 + x_2^2 + x^3 = c.$$

The solution set here is

- an **ellipsoid** if $c > 0$,



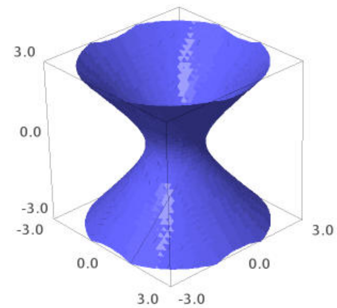
- a **point** if $c = 0$,
- empty if $c < 0$.

- $\text{sig}(Q) = (2, 1, 0)$. The representative equation is given by

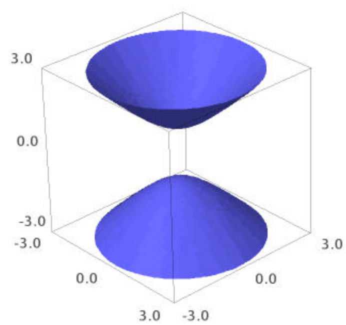
$$x_1^2 + x_2^2 - x_3^2 = c.$$

The solution set here is

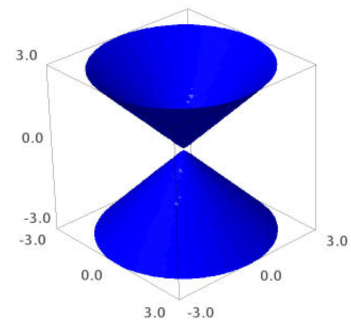
- a **hyperboloid of one sheet** if $c > 0$,



– a **hyperboloid of two sheets** if $c < 0$,



– a **double cone** if $c = 0$.

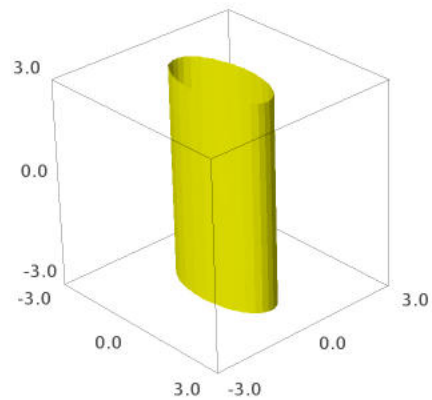


- $\text{sig}(Q) = (2, 0, 1)$. The representative equation is given by

$$x_1^2 + x_2^2 = c.$$

  The solution set here is

  - an **elliptic cylinder** if $c > 0$,



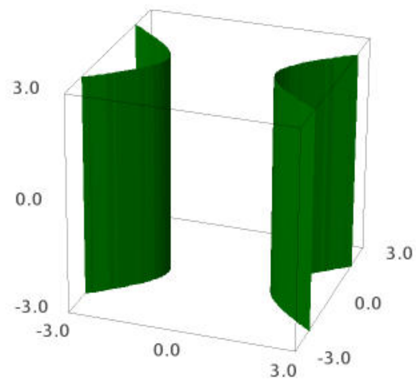
  - a **line** if $c = 0$,
  - empty if $c < 0$.

- $\text{sig}(Q) = (1, 1, 1)$. The representative equation is given by

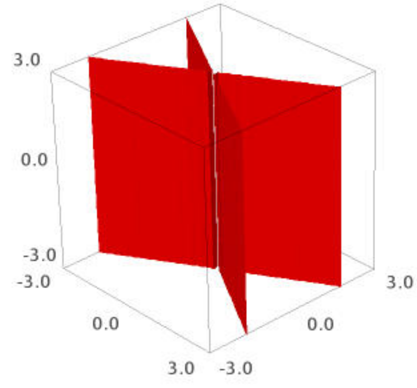
$$x_1^2 - x_2^2 = c.$$

  The solution set is

  - a **hyperbolic cylinder** if $c \neq 0$,



– a **pair of intersecting planes if $c = 0$.**



- $\text{sig}(Q) = (1, 0, 2)$. The representative equation is given by

$$x_1^2 = c.$$

  The solution set here is

  - a **pair of parallel planes** if $c > 0$,
  - a **single plane** if $c = 0$,
  - empty if $c < 0$.