Math 210A. Quadratic spaces over $\mathbf{R}$

1. Algebraic preliminaries

Let V be a finite free module over a nonzero commutative ring F. Recall that a *quadratic form* on V is a map $Q : V \to F$ such that $Q(cv) = c^2 Q(v)$ for all $v \in V$ and $c \in F$, and such that the symmetric pairing $V \times V \to F$ defined by $(v, w)_Q = Q(v + w) - Q(v) - Q(w)$ is *bilinear*. (The explicit coordinatized description, sometimes presented as the definition, will be given shortly.) A *quadratic space* over F is a pair (V, Q) consisting of a vector space V over F and a quadratic form Q on V.

Note that $(v, v)_Q = Q(2v) - 2Q(v) = 2Q(v)$, so as long as $2 \in F^\times$ (i.e., F is a $\mathbf{Z}[1/2]$-algebra) we can run the procedure in reverse: for any symmetric bilinear pairing $B : V \times V \to F$, $Q_B(v) = B(v, v)$ is a quadratic form on V and the two operations $Q \mapsto B_Q := (\cdot, \cdot)_Q/2$ and $B \mapsto Q_B$ are inverse bijections between quadratic forms on V and symmetric bilinear forms on V. Over general rings, one cannot recover Q from $(\cdot, \cdot)_Q$. (Example: $q(x) = x^2$ and $Q(x) = 0$ on $V = F$ have $(\cdot, \cdot)_q = 0 = (\cdot, \cdot)_Q$ when $2 = 0$ in F, yet $q \neq 0$.)

When $2 \in F^\times$, we say that Q is *non-degenerate* exactly when the associated symmetric bilinear pairing $(\cdot, \cdot)_Q : V \times V \to F$ is perfect (that is, the associated self-dual linear map $V \to V^\vee$ defined by $v \mapsto (v, \cdot)_Q = (\cdot, v)_Q$ is an isomorphism, or more concretely the "matrix" of $(\cdot, \cdot)_Q$ with respect to a basis of V is invertible). In other cases (still with $2 \in F^\times$) we say Q is *degenerate*. (There is a definition of non-degeneracy without assuming $2 \in F^\times$, but it is best to give it in terms of algebraic geometry.)

It is traditional in cases with $2 \in F^\times$ in F to put more emphasis on the symmetric bilinear form $B_Q = (\cdot, \cdot)_Q/2$ rather than on the symmetric bilinear form $(\cdot, \cdot)_Q$ (that is meaningful even if $2 \notin F$). Since we are not aiming to develop the general algebraic theory of quadratic forms over all rings or fields, and our main applications of the theory shall be over $\mathbf{R}$, we will generally restrict our attention to the case when $2 \in F^\times$.

If V has rank $n > 0$ we choose a basis $\{e_1, \dots, e_n\}$ of V, then for $v = \sum x_i e_i$ we have

$$Q(v) = Q(\sum_{i<n} x_i e_i + x_n e_n) = Q(\sum_{i<n} x_i e_i) + Q(x_n e_n) + (\sum_{i<n} x_i e_i, x_n e_n)_Q,$$

and bilinearity gives the last term as $\sum_{i<n} c_{in} x_i x_n$ with $c_{in} = (e_i, e_n)_Q \in F$. Also, $Q(x_n e_n) = c_{nn} x_n^2$ with $c_{nn} = Q(e_n) \in F$. Hence, inducting on the number of terms in the sum readily gives

$$Q(\sum x_i e_i) = \sum_{i \leq j} c_{ij} x_i x_j$$

with $c_{ij} \in F$, and conversely any such formula is readily checked to define a quadratic form. Note also that the c_{ij}'s are uniquely determined by Q (and the choice of basis): the formula forces $Q(e_i) = c_{ii}$, and then setting $x_i = x_j = 1$ for some $i < j$ and setting all other $x_k = 0$ gives $Q(e_i + e_j) = c_{ij} + c_{ii} + c_{jj}$, so indeed c_{ij} is uniquely determined. One could therefore say that a quadratic form "is" a homogeneous quadratic polynomial in the linear coordinates x_i's, but this coordinatization tends to hide underlying structure and make things seem more complicated than necessary, much like in the study of "matrix algebra" without the benefit of the theory of vector spaces and linear maps.

Example 1.1. Suppose $2 \in F^\times$, so we have seen that there is a bijective correspondence between symmetric bilinear forms on V and quadratic forms on V; this bijection is even linear with respect to the evident linear structures on the sets of symmetric bilinear forms on V and quadratic forms on V (using pointwise operations; $(a_1 B_1 + a_2 B_2)(v, v') = a_1 B_1(v, v') + a_2 B_2(v, v')$, which one checks is symmetric bilinear, and $(a_1Q_1 + a_2Q_2)(v) = a_1Q_1(v) + a_2Q_2(v)$ which as a function from V to F is checked to be a quadratic form). Let us make this bijection concrete, as follows. In class we saw that if we fix an ordered basis $\mathbf{e} = \{e_1, \dots, e_n\}$ of V then we can describe a symmetric bilinear $B : V \times V \to F$ in terms of the matrix $[B] = {}_{\mathbf{e}^\vee}[\varphi_\ell]_{\mathbf{e}} = (b_{ij})$ for the "left/right-pairing" map $\varphi_\ell = \varphi_r$ from V to $V^\vee$ defined by $v \mapsto B(v, \cdot) = B(\cdot, v)$, namely $b_{ij} = B(e_j, e_i) = B(e_i, e_j)$. However, in terms of the dual linear coordinates $\{x_i = e_i^*\}$ we have just seen that we can uniquely write $Q_B : V \to F$ as $Q_B(v) = \sum_{i \le j} c_{ij} x_i(v) x_j(v)$. What is the relationship between the c_{ij}'s and the b_{ij}'s?

We simply compute: for $v = \sum x_i e_i$, bilinearity of B implies that $Q_B(v) = B(v, v)$ is given by

$$\sum x_i x_j B(e_i, e_j) = \sum_i B(e_i, e_i) x_i^2 + \sum_{i<j} (B(e_i, e_j) + B(e_j, e_i)) x_i x_j = \sum_i b_{ii} x_i^2 + \sum_{i<j} 2 b_{ij} x_i x_j,$$

where $b_{ij} = B(e_j, e_i) = B(e_i, e_j) = b_{ji}$. Hence, $c_{ii} = b_{ii}$ and for $i < j$ we have $c_{ij} = 2b_{ij}$. Thus, for B and Q that correspond to each other, given the polynomial $[Q]$ for Q with respect to a choice of basis of V, we "read off' the symmetric matrix $[B]$ describing B (in the same linear coordinate system) as follows: the ii-diagonal entry of $[B]$ is the coefficient of the square term x_i^2 in Q, and the "off-diagonal" matrix entry b_{ij} for $i \ne j$ is given by half the coefficient for $x_i x_j = x_j x_i$ appearing in $[Q]$ (recall $2 \in F^\times$). For example, if $Q(x, y, z) = x^2 + 7y^2 - 3z^2 + 4xy + 3xz - 5yz$ then the corresponding symmetric bilinear form B is computed via the symmetric matrix

$$[B] = \begin{pmatrix} 1 & 2 & 3/2 \\ 2 & 7 & -5/2 \\ 3/2 & -5/2 & -3 \end{pmatrix}.$$

Going in the other direction, if someone hands us a *symmetric* matrix $[B] = (b_{ij})$ then we "add across the main diagonal" to compute that the corresponding homogeneous quadratic polynomial $[Q]$ is $\sum_i b_{ii} x_i^2 + \sum_{i<j} (b_{ij} + b_{ji}) x_i x_j = \sum_i b_{ii} x_i^2 + \sum_{i<j} 2 b_{ij} x_i x_j$.

It is an elementary algebraic fact (to be proved in a moment) for any field F with $\mathrm{char}(F) \ne 2$ there is a basis $\mathbf{e} = \{e_1, \dots, e_n\}$ of V with respect to which Q has the form $Q = \sum \lambda_i x_i^2$ for some scalars $\lambda_1, \dots, \lambda_n$ (some of which may vanish). In other words, we can "diagonalize" Q, or rather the "matrix" of B_Q (and so the property that some λ_i vanishes is equivalent to the intrinsic property that Q is degenerate). To see why this is, we note that Q is uniquely determined by B_Q (as $2 \in F^\times$ in F) and in terms of B_Q this says that the basis consists of vectors $\{e_1, \dots, e_n\}$ that are mutually perpendicular with respect to B_Q (i.e., $B_Q(e_i, e_j) = 0$ for all $i \ne j$). Thus, we can restate the assertion as the general claim that if $B : V \times V \to F$ is a symmetric bilinear pairing then there exists a basis $\{e_i\}$ of V such that $B(e_i, e_j) = 0$ for all $i \ne j$. To prove this we may induct on the rank, the case of rank 1 being clear. In general, if V has rank $n > 1$ then we can assume $B \ne 0$ and in this case we claim that there exists a nonzero $e_n \in V$ such that $B(e_n, e_n) \ne 0$. Granting this, $v \mapsto B(e_n, v) = B(v, e_n)$ is a nonzero linear functional on V whose kernel must be a hyperplane H that does not contain e_n, so by induction applied to B restricted to $H \times H$ we may find a suitable $e_1, \dots, e_{n-1}$ that, together with e_n, solves the problem.

It remains to show that for a nonzero symmetric bilinear form B on a finite-dimensional vector space V over a field F in which $2 \ne 0$, there must exist a nonzero $v_0 \in V$ such that $B(v_0, v_0) \ne 0$. We can reconstruct B from the quadratic form $Q_B(v) = B(v, v)$ via the formula

$$B(v, w) = \frac{Q_B(v + w, v + w) - Q_B(v, v) - Q_B(w, w)}{2}$$

since $2 \ne 0$ in F, so if $Q_B = 0$ then $B = 0$. This contradiction forces $Q_B(v_0) \ne 0$ for some $v_0 \in V$ if $B \ne 0$, as desired.

2. Some generalities over $\mathbf{R}$

Now assume that $F = \mathbf{R}$. Since all positive elements of $\mathbf{R}$ are squares, after first passing to a basis of V that "diagonalizes" Q (which, as we have seen, is a purely algebraic fact), we can rescale the basis vectors using $e_i' = e_i/\sqrt{|\lambda_i|}$ when $\lambda_i \neq 0$ to get (upon reordering the basis)

$$Q = {x'}_1^2 + \cdots + {x'}_r^2 - {x'}_{r+1}^2 - \cdots - {x'}_{r+s}^2$$

for some $r, s \geq 0$ with $r + s \leq \dim V$. Let $t = \dim V - r - s \geq 0$ denote the number of "missing variables" in such a diagonalization (so $t = 0$ if and only Q is non-degenerate). The value of r here is just the number of λ_i's which were positive, s is the number of λ_i's which were negative, and t is the number of λ_i's which vanish. The values r, s, t a priori may seem to depend on the original choice of ordered basis $\{e_1, \dots, e_n\}$.

To shed some light on the situation, we introduce some terminology that is specific to the case of the field $\mathbf{R}$. The quadratic form Q is *positive-definite* if $Q(v) > 0$ for all $v \in V - \{0\}$, and Q is *negative-definite* if $Q(v) < 0$ for all $v \in V - \{0\}$. Since $Q(v) = B_Q(v, v)$ for all $v \in V$, clearly if Q is either positive-definite or negative-definite then Q is non-degenerate. In terms of the diagonalization with all coefficients equal to ± 1 or 0, positive-definiteness is equivalent to the condition $r = n$ (and so this possibility is coordinate-independent), and likewise negative-definiteness is equivalent to the condition $s = n$. In general we define the *null cone* to be

$$C = \{v \in V \,|\, Q(v) = 0\},$$

so for example if $V = \mathbf{R}^3$ and $Q(x, y, z) = x^2 + y^2 - z^2$ then the null cone consists of vectors $(x, y, \pm\sqrt{x^2 + y^2})$ and this is physically a cone (or really two cones with a common vertex at the origin and common central axis). In general C is stable under scaling and so if it is not the origin then it is a (generally infinite) union of lines through the origin; for $\mathbf{R}^2$ and $Q(x, y) = x^2 - y^2$ it is a union of two lines.

Any vector v not in the null cone satisfies exactly one of the two possibilities $Q(v) > 0$ or $Q(v) < 0$, and we correspondingly say (following Einstein) that v is *space-like* or *time-like* (with respect to Q). The set V^+ of space-like vectors is an open subset of V, as is the set V^- of time-like vectors. These open subsets are disjoint and cover the complement of the null cone. In the preceding example with $Q(x, y) = x^2 - y^2$ on $V = \mathbf{R}^2$, V^+ and V^- are each disconnected (as drawing a picture shows quite clearly). This is atypical:

Lemma 2.1. *The open set V^+ in V is non-empty and path-connected if $r > 1$, with r as above in terms of a diagonalizing basis for Q, and similarly for V^- if $s > 1$.*

Proof. By replacing Q with $-Q$ if necessary, we may focus on V^+. Obviously V^+ if non-empty if and only if $r > 0$, so we may now assume $r \geq 1$. We have

$$Q(x_1, \dots, x_n) = x_1^2 + \cdots + x_r^2 - x_{r+1}^2 - \cdots - x_{r+s}^2$$

with $r \geq 1$ and $0 \leq s \leq n - r$. Choose $v, v' \in V^+$, so $x_j(v) \neq 0$ for some $1 \leq j \leq r$. We may move along a line segment contained in V^+ to decrease all $x_j(v)$ to 0 for $j > r$ (check!), and similarly for v', so for the purposes of proving connectivity we can assume $x_j(v) = x_j(v') = 0$ for all $j > r$. If $r > 1$ then v and v' lie in the subspace $W = \mathrm{span}(e_1, \dots, e_r)$ of dimension $r > 1$ on which Q has positive-definite restriction. Hence, $W - \{0\} \subseteq V^+$, and $W - \{0\}$ is path-connected since $\dim W > 1$. ■

The basis giving such a diagonal form is simply a basis consisting of r space-like vectors, s time-like vectors, and $n - (r + s)$ vectors on the null cone such that all n vectors are B_Q-perpencidular to each other. In general such a basis is rather non-unique, and even the subspaces

$$V_{+,\mathbf{e}} = \mathrm{span}(e_i \,|\, \lambda_i > 0), \quad V_{-,\mathbf{e}} = \mathrm{span}(e_i \,|\, \lambda_i < 0)$$

are *not* intrinsic. For example, if $V = \mathbf{R}^2$ and $Q(x,y) = x^2 - y^2$ then we can take $\{e_1, e_2\}$ to be either $\{(1,0),(0,1)\}$ or $\{(2,1),(1,2)\}$, and we thereby get different spanning lines. Remarkably, it turns out that the values

$$r_{\mathbf{e}} = |\{i \,|\, \lambda_i > 0\}| = \dim V_{+,\mathbf{e}}, \;\; s_{\mathbf{e}} = |\{i \,|\, \lambda_i < 0\}| = \dim V_{-,\mathbf{e}}, \;\; t_{\mathbf{e}} = |\{i \,|\, \lambda_i = 0\}| = \dim V - r_{\mathbf{e}} - s_{\mathbf{e}}$$

are independent of the choice of "diagonalizing basis" $\mathbf{e}$ for Q. One thing that is clear right away is that the subspace

$$V_{0,\mathbf{e}} = \mathrm{span}(e_i \,|\, \lambda_i = 0)$$

is actually intrinsic to V and Q: it is the set of $v \in V$ that are B_Q-perpendicular to the entirety of V: $B_Q(v, \cdot) = 0$ in $V^\vee$. (Beware that this is not the set of $v \in V$ such that $Q(v) = 0$; this latter set is the null cone C, and it is never a linear subspace of V when it contains nonzero points.)

3. Algebraic proof of well-definedness of the signature

Theorem 3.1. *Let V be a finite-dimensional $\mathbf{R}$ vector space, and Q a quadratic form on V. Let $\mathbf{e}$ be a diagonalizing basis for Q on V. The quantities $\dim V_{+,\mathbf{e}}$ and $dim V_{-,\mathbf{e}}$ are independent of $\mathbf{e}$.*

We'll prove this theorem using algebraic methods in a moment (and a longer, but more illuminating, proof by connectivity considerations will be given in the next section). In view of the intrinsic nature of the number of positive coefficients and negative coefficients in a diagonal form for Q (even though the specific basis giving rise to such a diagonal form is highly non-unique), we are motivated to make the:

Definition 3.2. Let Q be a quadratic form on a finite-dimensional $\mathbf{R}$-vector space V. We define the *signature* of (V, Q) (or of Q) to be the ordered pair of non-negative integers (r, s) where $r = \dim V_{+,\mathbf{e}}$ and $s = \dim V_{-,\mathbf{e}}$ respectively denote the number of positive and negative coefficients for a diagonal form of Q. In particular, $r + s \leq \dim V$ with equality if and only if Q is non-degenerate.

The signature is an invariant that is intrinsically attached to the finite-dimensional quadratic space (V, Q) over $\mathbf{R}$. In the study of quadratic spaces over $\mathbf{R}$ with the fixed dimension, it is really the "only" invariant. Indeed, we have:

Corollary 3.3. *Let (V, Q) and (V', Q') be finite-dimensional quadratic spaces over $\mathbf{R}$ with the same finite positive dimension. The signatures coincide if and only if the quadratic spaces are isomorphic; i.e., if and only if there exists a linear isomorphism $T : V \simeq V'$ with $Q'(T(v)) = Q(v)$ for all $v \in V$.*

This corollary makes precise the fact that the signature and dimension are the only isomorphism class invariants in the *algebraic* classification of finite-dimensional quadratic spaces over $\mathbf{R}$. However, even when the signature is fixed, there is a lot more to do than mere algebraic classification. There's a lot of geometry in the study of quadratic spaces over $\mathbf{R}$, so the algebraic classification via the signature is not the end of the story. We now prove the corollary, granting Theorem 3.1, and then we will prove Theorem 3.1.

Proof. Assume such a T exists. If $\mathbf{e}$ is a digonalizing basis for Q, clearly $\{T(e_i)\}$ is a diagonalizing basis for Q' with the same diagonal coefficients, whence Q' has the same signature as Q. Conversely, assume Q and Q' have the same signatures (r, s), so there exist ordered bases $\mathbf{e}$ and $\mathbf{e}'$ of V and

V' such that in terms of the corresponding linear coordinate systems $x_1, \dots, x_n$ and $x'_1, \dots, x'_n$ we have

$$Q = x_1^2 + \cdots + x_r^2 - x_{r+1}^2 - \cdots - x_{r+s}^2, \;\; Q' = x'^2_1 + \cdots + x'^2_r - x'^2_{r+1} - \cdots - x'^2_{r+s}.$$

Note in particular that

$$Q(\sum a_i e_i) = \sum_{i=1}^{r} a_i^2 - \sum_{i=r+1}^{s} a_i^2 = Q'(\sum a_i e'_i)$$

for all i. Thus, if $T : V \to V'$ is the linear map determined by $T(e_i) = e'_i$ then T sends a basis to a basis. Thus, T is a linear isomorphism, and also

$$Q'(T(\sum a_i e_i)) = Q'(\sum a_i e'_i) = Q(\sum a_i e_i).$$

In other words, $Q' \circ T = Q$, as desired. ■

Now we turn to the proof of the main theorem stated above.

Proof. Let $V_0 = \{v \in V \,|\, B_Q(v, \cdot) = 0\}$. Let $\mathbf{e} = \{e_1, \dots\}$ be a diagonalizing basis of V for Q, with $Q = \sum \lambda_i x_i^2$ relative to $\mathbf{e}$-coordinates, where $\lambda_1, \dots, \lambda_{r_\mathbf{e}} > 0$, $\lambda_{r_\mathbf{e}+1}, \dots, \lambda_{r_\mathbf{e}+s_\mathbf{e}} < 0$, and $\lambda_i = 0$ for $i > r_\mathbf{e} + s_\mathbf{e}$. Clearly $V_0 = V_{0,\mathbf{e}}$.

Now we consider the subspaces $V_{+,\mathbf{e}}$ and $V_{-,\mathbf{e}}$. Since these subspaces (along with $V_0 = V_{0,\mathbf{e}}$) are given as the span of parts of the basis $\mathbf{e}$ (chopped up into three disjoint pieces, some of which may be empty), we have a decomposition

$$V = V_{+,\mathbf{e}} \oplus V_{-,\mathbf{e}} \oplus V_0$$

with $B_Q(v^+, v^-) = 0$ for all $v^+ \in V_{+,\mathbf{e}}$, $v^- \in V_{-,\mathbf{e}}$ (due to the diagonal shape of B_Q relative to the $\mathbf{e}$-coordinates). Consider the quotient space V/V_0 that has nothing to do with $\mathbf{e}$. The bilinear form B_Q and the quadratic form Q are *well-defined* on V/V_0, and the original dimensions for $V_{\pm,\mathbf{e}}$ coincide with the dimensions of the analogous subspaces of V/V_0 using the induced quadratic form (via Q) and induced basis arising from the part of $\mathbf{e}$ not in V_0. In this way we are reduced to considering the case $V_0 = 0$, which is to say that Q is non-degenerate on V.

It is clear from the diagonal form of Q that $Q(v^+) \geq 0$ for all $v^+ \in V_{+,\mathbf{e}}$ with $Q(v^+) = 0$ if and only if $v^+ = 0$, since Q on the subspace $V_+ = V_{+,\mathbf{e}}$ is presented as the sum of squares of basis coordinates. Likewise, we have $Q(v^-) \leq 0$ for $v^- \in V_- = V_{-,\mathbf{e}}$, with equality if and only if $v^- = 0$. Hence, if we consider another diagonalizing basis $\mathbf{e}'$ and the resulting decomposition $V'_+ \oplus V'_-$ of V then $V_+ \cap V'_- = 0$ and $V_- \cap V'_+ = 0$ since a vector in either overlap has both non-negative and non-positive (hence vanishing) value under the quadratic form Q that is definite on each of the spaces $V_\pm$ and $V'_\pm$.

The vanishing of these overlaps gives

$$\dim V_+ + \dim V'_- \leq \dim V, \;\; \dim V_- + \dim V'_+ \leq \dim V,$$

but since $V_+ \oplus V_- = V = V'_+ \oplus V'_-$ we also have

$$\dim V_+ + \dim V_- = \dim V = \dim V'_+ + \dim V'_-.$$

Hence, $\dim V_+ \leq \dim V'_+$ and $\dim V_- \leq \dim V'_-$. Switching the roles of the two decompositions gives the reverse inequalities, so equalities are forced as desired. ■

4. Geometric proof of well-definedness of signature

We now provide an alternative geometric approach that gives an entirely different (and rather more interesting and vivid) proof that the signature is well-defined. The key geometric input will be the connectivity of $\mathrm{GL}^+(V)$ (the subgroup of linear automorphisms with positive determinant); such connectivity is left to the reader as an exercise with Gramm-Schmidt orthogonalization. The proof below is somewhat longer than the largely algebraic method used above, but it brings out the group-theoretic and topological structures that are lying in the shadows.

Let us fix a positive-definite inner product $\langle \cdot, \cdot \rangle$ on V. Every bilinear form B on V may therefore be expressed as $B(v, v') = \langle T(v), v' \rangle$ for a unique self-map $T : V \to V$, and symmetry (resp. non-degeneracy) of B is the condition that T be self-adjoint (resp. an isomorphism). Note that the formation of T depends on not only B but also on the choice of $\langle \cdot, \cdot \rangle$. Consider the self-adjoint map $T_Q : V \to V$ associated to B_Q and to the initial choice of inner product $\langle \cdot, \cdot \rangle$ on V. (That is, $B_Q(v, v') = \langle T_Q(v), v' \rangle$ for all $v, v' \in V$.) The condition that a basis $\mathbf{e} = \{e_i\}$ diagonalize Q is exactly the condition that $\langle T_Q(e_i), e_j \rangle = 0$ for all $i \neq j$. That is, this says that $T_Q(e_i)$ is perpendicular to e_j for all $j \neq i$. In particular, if $\mathbf{e}$ were an *orthogonal* (e.g., orthonormal) basis with respect to $\langle \cdot, \cdot \rangle$ then the diagonalizability condition would say that $\mathbf{e}$ is a basis of eigenvectors for T_Q. We can now run this procedure partly in reverse: if we start with a basis $\mathbf{e}$ that diagonalizes Q, then we can *define* an inner product $\langle \cdot, \cdot \rangle_{\mathbf{e}}$ by the condition that it makes $\mathbf{e}$ orthonormal, and the resulting self-adjoint $T_{Q,\mathbf{e}}$ then has its number of positive (resp. negative) eigenvalues given by $r_{\mathbf{e}}$ and $s_{\mathbf{e}}$ when these numbers of eigenvalues are counted with multiplicity (as roots of the characteristic polynomials of $T_{Q,\mathbf{e}}$).

We may now exploit the flexibility in the choice of the inner product to restate our problem in terms of arbitrary inner products on V rather than in terms of diagonalizing bases for Q: for each positive-definite inner product $I = \langle \cdot, \cdot \rangle$ on V we have $B_Q = \langle T_{Q,I}(\cdot), \cdot \rangle$ for a unique map $T_{Q,I} : V \to V$ that is self-adjoint with respect to I, and we let r_I and s_I denote the respective number of positive and negative eigenvalues of $T_{Q,I}$ (with multiplicity). Here the spectral theorem enters: it ensures that for *any* choice of I, $T_{Q,I}$ does diagonalize over $\mathbf{R}$. Our problem can therefore be recast as that of proving that r_I and s_I are independent of I. Roughly speaking, to each I we have attached a pair of discrete (i.e., $\mathbf{Z}$-valued) parameters r_I and s_I (using Q), and so if the "space" of I's is connected in a reasonable sense then discrete parameters on this space should not jump. That is, if we can topologize the space of I's such that r_I and s_I depend continuously on I then connectivity of such a topology would give the desired result.

The existence of an orthonormal basis for any I, coupled with the fact that $\mathrm{GL}(V)$ acts transitively on the set of ordered bases of V (i.e., for any two ordered bases $\{e_1, \cdots, e_n\}$ and $\{e'_1, \ldots, e'_n\}$ there exists a (unique) linear automorphism L of V such that $L(e_i) = e'_i$ for all i), implies that $\mathrm{GL}(V)$ acts transitively on the set of I's. That is, if $I = \langle \cdot, \cdot \rangle$ and $I' = \langle \cdot, \cdot \rangle'$ are two inner products on V then there exists $L \in \mathrm{GL}(V)$ such that $\langle v, v' \rangle = \langle L(v), L(v') \rangle'$. Concretely, L carries an ordered orthonormal basis with respect to I to one with respect to I'. This shows slightly more: at the expense of replacing one of the ONB vectors with its negative we can flip the sign of $\det L$. Hence, even the *connected* $\mathrm{GL}^+(V)$ acts transitively on the set of all I's. This leads to:

Theorem 4.1. *Let W be the finite-dimensional vector space of symmetric bilinear forms on V, endowed with its natural topology as a finite-dimensional vector space over $\mathbf{R}$. The subset of elements that are positive-definite inner products is open and connected.*

Proof. We first prove connectedness, and then we prove openness. There is a natural left action of $\mathrm{GL}(V)$ on W: to $L \in \mathrm{GL}(W)$ and $B \in W$, we associate the symmetric bilinear form $L.B = B(L^{-1}(\cdot), L^{-1}(\cdot))$. By fixing a basis of V and computing in linear coordinates we see that the resulting map

$$\mathrm{GL}(V) \times W \to W$$

is continuous. In particular, if we fix $B_0 \in W$ then the map $\mathrm{GL}(V) \to W$ defined by $L \mapsto L.B_0$ is continuous. Restricting to the connected subgroup $\mathrm{GL}^+(V)$, it follows from continuity that the $\mathrm{GL}^+(V)$-orbit of any B_0 is connected in W. But if we take B_0 to be an inner product then from the definition of the action we see that $L.B_0$ is an inner product for every $L \in \mathrm{GL}^+(V)$ (even for $L \in \mathrm{GL}(V)$), and it was explained above that *every* inner product on V is obtained from a single B_0 by means of some $L \in \mathrm{GL}^+(V)$. This gives the connectivity.

Now we check openness. This says that the "positive-definiteness" property of a symmetric bilinear form cannot be lost under small deformation. Fix an inner product $\langle \cdot, \cdot \rangle_0$ on V, and let S_0 be the resulting compact unit sphere. For any symmetric bilinear form B on V, it is clear that B is positive definite if and only if the function $Q_B = B(v, v)$ restricted to the compact S_0 has positive lower bound. By compactness it is obvious that for any B' sufficiently close to B in the sense of the natural topology on the linear space of symmetric bilinear forms, the lower bound for $Q_{B'}|_{S_0}$ is near to that of $Q_B|_{S_0}$, and so indeed B' is positive-definite for B' near B. ■

We have now finished the proof of Theorem 4.1, so the space of inner products I on V has been endowed with a natural connected topology, and it remains to show that the $\mathbf{Z}$-valued functions $I \mapsto r_I$ and $I \mapsto s_I$ that count the number of positive (resp. negative) roots of $T_{Q,I}$ (with multiplicity!) are continuous in I. Put another way, the dependence on I is locally constant: if I' is sufficiently close to I then we claim that $r_{I'} = r_I$ and $s_{I'} = s_I$. If we let χ_I denote the characteristic polynomial of $T_{Q,I}$, then the number of zeros of $\chi_I(z)$ is independent of I: it is exactly the dimension $t = \dim V_0$ of the space of $v \in V$ such that $B_Q(v, \cdot) = 0$. Hence, the polynomials $\chi_I(z)/z^t \in \mathbf{C}[z]$ have all roots in $\mathbf{R}^\times$, and our problem is to study the variation in the number r_I of positive roots of this latter polynomial (this determines the number of negative roots, $s_I = n - t - r_I$) as we slightly move I. To proceed, we need to prove a lemma that is usually called "continuity of roots":

Lemma 4.2. *Let $f = z^n + c_{n-1}z^{n-1} + \cdots + c_0 \in \mathbf{C}[z]$ be a monic polynomial with positive degree n, and let $\{z_i\}$ be the set of distinct roots of f in $\mathbf{C}$. For any $\varepsilon > 0$ there exists $\delta > 0$ such that if $g = z^n + b_{n-1}z^{n-1} + \cdots + b_0 \in \mathbf{C}[z]$ is monic of degree n with $|b_j - c_j| < \delta$ for all $j < n$ then each root ρ of g in $\mathbf{C}$ satisfies $|\rho - z_i| < \varepsilon$ for some i.*

Moreover, if $\varepsilon < \min_{i \neq i'} |z_i - z_{i'}|/2$ and μ_i is the multiplicity of z_i as a root of f (so $\sum \mu_i = n$) then by taking δ to be sufficiently small there are exactly μ_i roots ρ of g – counting with multiplicity – such that $|\rho - z_i| < \delta$.

The astute reader will check that the proof of the lemma works if we replace $\mathbf{C}$ with $\mathbf{R}$ throughout (which suffices for the intended applications). However, the lemma is rather much weaker when stated over $\mathbf{R}$, due to the general lack of real roots to polynomials over $\mathbf{R}$.

Proof. We first fix any $\varepsilon > 0$ and prove the existence of δ as in the first assertion in the lemma. Assume to the contrary that no such δ exists, so let $g_m = z^n + b_{n-1,m}z^{n-1} + \cdots + b_{0,m}$ satisfy $b_{j,m} \to c_j$ for all $j < n$ such that there exists a root $\rho_m \in \mathbf{C}$ of g_m such that $|\rho_m - z_i| \geq \varepsilon$ for all i. By elementary upper bounds on roots of *monic* polynomials in terms of lower-degree coefficients (and the degree of the polynomial), since the $|b_{j,m}|$'s are bounded it follows that the $|\rho_m|$'s are bounded. Hence, by compactness of closed discs in $\mathbf{C}$ we may pass to a subsequence of the g_m to arrange that $\{\rho_m\}$ has a limit $\rho \in \mathbf{C}$, and by passing to the limit $|\rho - z_i| \geq \varepsilon$ for all i. However, $b_{j,m} \to c_j$ for all $j < n$, so $0 = g_m(\rho_m) \to f(\rho)$. This contradicts the fact that ρ is distinct from all of the roots z_i of f in $\mathbf{C}$.

Now take ε smaller than half the minimum distance between distinct roots of f, so by taking δ sufficiently small (in accordance with ε) each root ρ of g satisfies $|\rho - z_i| < \varepsilon$ for a *unique* root z_i of f when the coefficients of g satisfy $|b_j - c_j| < \delta$ for all $j < n$. This uniqueness of z_i for each ρ is due to the smallness of ε. In this way, we have a map from the set of roots of g to the set of roots of f, assigning to each root ρ of g the unique root of f to which it is closest. We want to prove that by taking δ sufficiently small, exactly μ_i roots of g (with multiplicity) are closest (even within a distance $< \varepsilon$) to the root z_i of f. Assuming no such δ exists, since there are only finitely many z_i's we may use a pigeonhole argument (and relabelling of the z_i's) to make a sequence of g_m's with $b_{j,m} \to c_j$ such that the number of roots of g_m within a distance $< \varepsilon$ from z_1 is equal to a fixed non-negative integer $\mu \neq \mu_1$. Consider a monic factorization

$$g_m(z) = \prod_{j=1}^{n} (z - \rho_{j,m})$$

with $|\rho_{j,m} - z_{i(j)}| < \varepsilon$ for a unique $i(j)$ for each m. There are exactly μ values of j such that $i(j) = 1$.

By the same compactness argument as above, we can pass to a subsequence of the g_m's so that $\{\rho_{j,m}\}_{m\geq 1}$ has a limit ρ_j satisfying $|\rho_j - z_{i(j)}| \leq \varepsilon$. Due to the smallness of ε, $z_{i(j)}$ is the unique root of f that is so close to ρ_j. In particular, there are μ values of j for which ρ_j is closer to z_1 than to any other roots of f, and for all other j the limit ρ_j is closer to some other root of f than it is to z_1. However, since $g_m \to f$ coefficient-wise it follows that $f(z) = \prod_{j=1}^{n}(z - \rho_j)$. Hence, there are exactly μ_1 values of j such that $\rho_j = z_1$ and for all other values of j we have that ρ_j is equal to z_i for some $i \neq 1$. This contradicts the condition $\mu \neq \mu_1$. ■

By the lemma on continuity of roots (applied with $f = \chi_I(z)/z^t$ and $g = \chi_{I'}/z^t$ for I' near I), our problem is reduced to proving that $\chi_{I'}$ is coefficient-wise close to χ_I for I' near to I in the space of inner products on V. Such closeness would follow from $T_{Q,I'}$ being sufficiently close to $T_{Q,I}$ in $\mathrm{Hom}(V, V)$, so we are reduced to proving that by taking I' sufficiently close to I we make $T_{Q,I'}$ as close as we please to $T_{Q,I}$. If $L : V \simeq V$ is a linear isomorphism carrying I to I' (i.e., $\langle L(v), L(v')\rangle = \langle v, v'\rangle'$) then

$$\langle T_{Q,I}(v), v'\rangle = B_Q(v, v') = \langle T_{Q,I'}(v), v'\rangle' = \langle L(T_{Q,I'}(v)), L(v')\rangle = \langle (L^*L \circ T_{Q,I'})(v), v'\rangle,$$

where L^* is the I-adjoint of L, so $T_{Q,I'} = L^*LT_{Q,I}$. Note that the initial condition on L only determines it up to left-multiplication by an element in the orthogonal group of I, and this ambiguity cancels out in L^*L. Hence, L^*L is well-defined in terms of I' and I. In particular, if we consider I as fixed and I' as varying then L^*L is a $\mathrm{GL}(V)$-valued function of I', and our problem is reduced to proving that for I' sufficiently near I we have $(L^*L)^{-1}$ sufficiently near the identity (as this makes $T_{Q,I'} = (L^*L)^{-1}T_{Q,I}$ sufficiently near $T_{Q,I}$, where "sufficiently near" of course depends on I and more specifically on $T_{Q,I}$).

The identity

$$\langle v, v'\rangle' = \langle L(v), L(v')\rangle = \langle (L^*L)(v), v'\rangle$$

implies that if we fix a basis $\mathbf{v}$ of V and let M and M' be the associated invertible symmetric matrices computing $\langle \cdot, \cdot\rangle$ and $\langle \cdot, \cdot\rangle'$ then $M' = (L^*L)M$ and the *definition* of the topology on the space of inner products says that $M' - M$ is very close to zero. Hence, we can restate the problem as proving that for a fixed invertible matrix M and any matrix M' sufficiently close to M (entry by entry, and so in particular M' is invertible as $\det(M')$ is near $\det(M) \neq 0$), the matrix $M(M')^{-1}$ is near the identity. Working in the language of sequences (which is to say, arguing by contradiction), we want to show that if $\{M_s\}$ is a sequence of invertible matrices with $M_s \to M$ then $MM_s^{-1} \to MM^{-1} = 1$.

This follows from the continuity of both matrix multiplication and Cramer's formula for the inverse of a matrix, and so completes the geometric proof of the well-definedness of the signature.

We now use the preceding geometric technique to prove a generalization of Theorem 4.1:

Corollary 4.3. *Let W be the finite-dimensional vector space of symmetric bilinear forms on V, endowed with its natural topology as a finite-dimensional vector space over $\mathbf{R}$. Let W^0 be the subset of non-degenerate symmetric bilinear forms. The subset W^0 is open in W and it has finitely many connected components: its connected components consist of those B's having a fixed signature (r,s) with $r+s=\dim V$.*

In the positive-definite case, this recovers Theorem 4.1.

Proof. In terms of the "matrix" description of points $B \in W$ with respect to a choice of ordered basis of V, B is non-degenerate if and only if its associated symmetric matrix (a_{ij}) has non-vanishing determinant. In other words, the subset $W^0 \subseteq W$ is the non-vanishing locus of a polynomial function in linear coordinates and so it is open. We now fix an ordered pair (r,s) of non-negative integers satisfying $r+s=\dim V$ and we let $W^0_{(r,s)}$ be the subset of points $B \in W^0$ whose associated quadratic form $Q_B : V \to \mathbf{R}$ has signature (r,s). Our goal is to prove that the subsets $W^0_{(r,s)}$ are the connected components of W^0. Note that since W^0 is open in a vector space, its connected components are open subsets.

We have to prove two things: the signature is locally constant on W^0 (and hence is constant on connected components of W^0), and each $W^0_{(r,s)}$ is connected. For connectivity, we may use the exact same argument as in the beginning of the proof of Theorem 4.1 once we prove that any two quadratic forms $q, q' : V \to \mathbf{R}$ with the same signature (r,s) are related by the action of $\mathrm{GL}^+(V)$ on V. The quadratic spaces (V,q) and (V,q') are certainly isomorphic since q and q' have the same signature, so there exists $T \in \mathrm{GL}(V)$ such that $q' = q \circ T$. The only potential snag is that $\det T \in \mathbf{R}^\times$ might be negative. To fix this, we just need to find $T_0 \in \mathrm{GL}(V)$ such that $\det T_0 < 0$ and $q = q \circ T_0$, as then we could replace T with $T_0 \circ T \in \mathrm{GL}^+(V)$. To find T_0, we argue exactly as in the positive-definite case: we find an ordered basis $\mathbf{e} = \{e_1, \ldots, e_n\}$ of V with respect to which q is diagonalized, and we let $T_0 : V \simeq V$ be the map that negates e_1 but fixes e_j for all $j > 1$. (Check that indeed $q \circ T_0 = q$.)

It remains to show that if $B \in W^0$ is a point such that Q_B has signature (r,s), then for all $B' \in W^0$ near B the non-degenerate quadratic form $Q_{B'}$ on V also has signature (r,s). It is sufficient to track r, since $r+s=\dim V$. (Warning: It is crucial here that we assume B is non-degenerate. If $B \in W$ is a degenerate quadratic form, there are $B' \in W$ that are arbitrarily close to B and non-degenerate, so such B' have signature not equal to that of B. For a concrete example with $V = \mathbf{R}^2$, note that for small $\varepsilon > 0$

$$B_\varepsilon((x_1,x_2),(y_1,y_2)) = x_1y_1 - \varepsilon x_2y_2$$

in W^0 is very close to the degenerate $B_0 \in W$.)

We fix an inner product $\langle\cdot,\cdot\rangle$ on V and write $B = \langle T(\cdot), \cdot\rangle$ for a unique isomorphism $T : V \simeq V$ that is self-adjoint with respect to the inner product. The points $B' \in W$ have the form $B' = \langle T'(\cdot), \cdot\rangle$ for unique self-adjoint linear maps $T' : V \simeq V$, and this identifies W with the subspace of self-adjoint elements in $\mathrm{Hom}(V,V)$; under this identification, W^0 corresponds to the self-adjoint automorphisms of V. The condition that B' be close to B in W is exactly the condition that T' be close to T in $\mathrm{Hom}(V,V)$ (as the linear isomorphism of W onto the subspace of self-adjoint elements in $\mathrm{Hom}(V,V)$ is certainly a homeomorphism, as is any linear isomorphism between finite-dimensional $\mathbf{R}$-vector spaces). Hence, our problem may be restated as this: we fix a self-adjoint isomorphism $T : V \simeq V$, and we seek to prove that any self-adjoint isomorphism $T' : V \simeq V$ sufficiently close to T (in $\mathrm{Hom}(V, V)$) has the same number of positive eigenvalues as T (counting with multiplicities). Consider the characteristic polynomials $\chi_T, \chi_{T'} \in \mathbf{R}[\Lambda]$. These are monic polynomials of the same degree $n > 0$, and each has all complex roots in $\mathbf{R}$ (by the spectral theorem). Making T' approach T has the effect of making $\chi_{T'}$ "approach" χ_T for coefficients in each fixed degree (from 0 to $n-1$). Lemma 4.2 therefore gives the desired result, since χ_T does not have zero as a root. ■