Math 395. Geometric approach to signature

For the amusement of the reader who knows a tiny bit about groups (enough to know the meaning of a transitive group action on a set), we now provide an alternative geometric approach that gives an entirely different (and rather more interesting and vivid) proof that the signature of a real quadratic space is well-defined. (Our initial proof was largely algebraic.) The key geometric input will be the result on connectivity of $\mathrm{GL}^+(V)$, which we proved using a dynamic interpretation of Gram–Schmidt. The proof is somewhat longer than the algebraic method, but it nicely brings out the group-theoretic and topological structures that are lying in the shadows.

1. Preliminary steps

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. (Of course, the spectral theorem ensures that the self-adjoint T_Q *can* be diagonalized, but this is not logically relevant here because we are beginning with the e_i's.) 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 the connectivity of such a topology will 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 1.1. *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 $Q_B = B(v, v)/2$ 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. ∎

2. A local constancy argument

We have now finished the proof of Theorem 1.1, 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 2.1. *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 $T_{Q,I'} = L^* L T_{Q,I}$. 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 1.1:

Corollary 2.2. *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 1.1.

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 1.1 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 $\mathbf{e} = \{e_1, \dots, e_n\}$ 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 2.1 therefore gives the desired result, since χ_T does not have zero as a root. ■