SOME NOTES ON COMPLEX ANGLES

K.M. HALPERN

September 23, 2025

Contents

1. Intro

We have an intuitive notion of what lines and angles and lengths mean in a real inner-product space. Even though we can visualize such things only in two or three dimensions, this visualization readily lends itself to their interpretation in higher dimensions.

However, our intuition for $\mathbb{C}^n$ tends to be far less developed. The best we can really do is view it as $\mathbb{R}^{2n}$, which limits us to $n = 1$ in terms of visualization. The notion of length translates to the complex case easily enough, and complex lines become real planes, but defining and interpreting a complex angle is less straightforward.

One reason that we care about complex angles is quantum mechanics. Although calculations are performed in a Hilbert space, the true state space is a projective Hilbert space. For real vector spaces, an inner product doesn't survive projectivization, but the associated notion of angle does. It is reasonable to expect similar behavior in the complex case. An angle therefore would constitute an important piece of structure on the state space of quantum mechanics. Such an angle can indeed be defined, and it induces a Riemannian

metric known as the "Fubini-Study metric" on the projective space. In fact, it helps impart to the projective space the structure of a Kahler manifold.

In these notes, we'll explore the notion of complex angle in some detail, reserving the development of complex manifolds and the Fubini-Study metric for another day. We begin with a review of linear algebra, complex and real inner products, and the decomposition of a complex inner product into a real inner product and symplectic form. We then consider a number of potential definitions for a complex angle and find that all but one exhibit critical shortcomings. This motivates the use of the surviving candidate as our definition of a "complex angle".

2. Inner Products, Hermitian Forms, and Symplectic Forms

2.1. **Basic Definitions.**

Let's recall some definitions from linear algebra. Let V be a vector space over field K, and let V^* denote its dual: the vector space of linear maps $V \to K$. Although V^* and V are isomorphic, there is no natural isomorphism between them absent additional structure or information. However, V^{**} and V are naturally isomorphic. We'll sometimes use $dd : V \to V^{**}$ (for "double dual") to denote the natural isomorphism. Any basis on V induces an isomorphism between V and V^* and thus a basis for V^*.

The natural isomorphism $V \to V^{**}$ is given by $v^{**}(\omega) \equiv \omega(v)$ for all $\omega \in V^*$.

A choice of basis constitutes "additional information". Given a basis $\{e_i\}$ on V, we have a canonical basis $\{e'_i\}$ on V^* defined via $e'_i(e_j) = \delta_{ij}$. If we start with $\{e'_i\}$ instead, the canonical basis is defined via $e_i^{**}(e'_j) = \delta_{ij}$, where e_i^{**} is the canonical double-dual to e_i as just described. The map $e_i \to e'_i$ is the canonical isomorphism induced by the choice of basis.

We haven't defined them yet, but an inner product or symplectic form also induce a canonical isomorphism. The inner product $g : V \times V \to K$ does so via $v \to g(v, -)$ (which equals $g(-, v)$ by symmetry of the inner product). I.e., $v^*(w) = g(v, w)$. A symplectic form $\omega : V \times V \to K$ does so via $v \to \omega(v, -)$ (or $-\omega(-, v)$, which is the same when K has characteristic not equal to 2, since ω is then antisymmetric). Both the inner product and symplectic form are nondegenerate, so these are both isomorphisms. In fact, any nondegenerate bilinear form on V induces two canonical isomorphisms between V and V^*, which are identical if the form is symmetric. As we will discuss, a nondegenerate sesquilinear form does *not* induce an isomorphism from V to V^*, but rather from V to the conjugate dual space. In the presence of an inner product, there is a preferred class of bases as well: the orthonormal bases, in which g is the identity matrix. In the presence of a symplectic form on K^{2n}, there also is a preferred class of bases: the symplectic (aka canonical) bases, in which ω is $\begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}$, where I_n denotes the $n \times n$ identity matrix.

An (n, m)-tensor $f : V^* \times \cdots \times V^* \times V \times \cdots \times V \to K$ (with n copies of V^* and m copies of V) is linear in each of its arguments and thus "multilinear". It is easy to show that the (n, m)-tensors form a vector space, and the tensors as a whole form a graded algebra under the tensor product. The fully antisymmetric $(0, n)$-tensors (over all n) form a graded vector subspace but not a subalgebra. However, this graded subspace can be endowed with a canonical multiplication of its own, and the resulting graded algebra is known as the "exterior algebra" on V.

In the presence of an inner product, this same graded subspace can be endowed with another canonical multiplication known as the "Clifford product", resulting in the (non-graded) "Clifford algebra". The exterior algebra is the Clifford algebra when the inner product is trivial (aka 0). Alternately, we can think of the Clifford algebra as modifying the exterior algebra to convert symmetries to traces (via the inner product) rather than nullifying them. We'll develop all these notions in detail in an upcoming set of notes on the angle between two subspaces of $\mathbb{R}^n$.

Just as a choice of basis on V determines one on V^*, it also induces one on every (n, m)-tensor space, as well as on the tensor and exterior algebras.

A $(0,1)$-tensor is a dual vector, and a $(1,0)$-tensor can be viewed as a vector via the natural isomorphism dd. A $(1,1)$-tensor can be thought of as a linear operator on V or a linear operator on V^* or a bilinear map $V^* \times V \to K$, and a $(0,2)$-tensor can be thought of as a linear map $V \to V^*$ in two distinct ways or as a bilinear map $V \times V \to K$.

Confusingly, a $(1,1)$-tensor is sometimes referred to as a "matrix". We'll reserve the term for an array of numbers rather than an abstract object that transforms a certain way under basis changes.

Let f be a $(1,1)$-tensor, so $f : V^* \times V \to K$. The linear map $f(-,v) : V^* \to K$ is an element of V^{**} and thus has a natural counterpart in V via dd^{-1}. We therefore get a linear map $L : V \to V$ defined as $L(v) = dd^{-1}(f(-,v))$. By similar reasoning, we have a linear map $L^* : V^* \to V^*$ defined as $L^*(\omega) = f(\omega, -)$.

For a $(0,2)$-tensor $f : V \times V \to K$, we can follow a similar approach. $f(-,v)$ is a linear map $V \to K$, and thus an element of V^*. We thus have a linear map $V \to V^*$ given by $v \to f(-,v)$. However, $v \to f(v,-)$ is another such linear map, so we have two ways of viewing f in this light. Similarly, a $(2,0)$-tensor $f : V^* \times V^* \to K$ can be viewed as a linear map $V^* \to V$ in two ways: $\omega \to dd^{-1}(f(\omega,-))$ or $\omega \to dd^{-1}(f(-,\omega))$.

The following are some common linear or linear-like objects on vector spaces:

- A **bilinear form** is just a $(0,2)$-tensor. I.e., $B : V \times V \to K$ is linear in each argument. As mentioned, it can also be viewed as a linear map $B : V \to V^*$ in two ways (depending which argument we fix). A bilinear form can be:
  - ◇ **symmetric**: $B(v,w) = B(w,v)$.
  - ◇ **antisymmetric** (aka **skew-symmetric**): $B(v,w) = -B(w,v)$.
  - ◇ **positive-definite**: $B(v,v) > 0$ for all $v \neq 0$.
  - ◇ **nondegenerate**: $B(v,w) = 0$ for all v iff $w = 0$, and ditto swapping v and w.

    It turns out that one implies the other. This is easiest to see by choosing a basis and representing the form as a matrix. Nondegeneracy is then equivalent to $\det M \neq 0$. Since $\det M = \det M^T$, one holds iff the other does.
- A **sesquilinear form** is like a bilinear form, but it is conjugate-linear in one of its arguments (which we'll take to be the second, though the first is commonly chosen as well). I.e., $B : V \times V \to K$ satisfies $B(av+v', w) = aB(v,w)+B(v',w)$ and $B(v, bw+w') = \bar{b}B(v,w)+B(v,w')$. A sesquilinear form can be:
  - ◇ **Hermitian**: $B(v,w) = \overline{B(w,v)}$. In that case $B(v,v)$ is always real.
  - ◇ **skew-Hermitian**: $B(v,w) = -\overline{B(w,v)}$. In that case, $B(v,v)$ is always imaginary.
  - ◇ **positive-definite**: $B(v,v)$ is real and > 0 for all $v \neq 0$.
  - ◇ **nondegenerate**: $B(v,w) = 0$ for all v iff $w = 0$, and ditto swapping v and w.

    As in the bilinear case, one implies the other.
- An **inner product** is a positive-definite Hermitian form. I.e., $I(v,w)$ is linear in the 1st argument, satisfies $I(v,w) = \overline{I(w,v)}$ (which implies conjugate-linearity in the 2nd argument), and $I(v,v)$ is real and positive for all $v \neq 0$.

  A word of warning: some mathematics books define an inner product to be nondegenerate rather than positive-definite.
- A **symplectic form** is a nondegenerate anti-symmetric bilinear form. I.e., $\omega(v,w) = -\omega(w,v)$, ω is linear in each argument, and $\omega(v,w) = 0$ for all v iff $w = 0$.

  Technically, when $charK = 2$ we must be careful to distinguish between antisymmetry and the condition $\omega(v,v) = 0$. We only care about $K = \mathbb{R}, \mathbb{C}$, both of which have $charK = 0$.

If $K = \mathbb{R}$, sesquilinear is the same as bilinear, Hermitian is the same as symmetric, anti-Hermitian is the same as antisymmetric, and an inner-product is just a positive-definite symmetric bilinear form. For $K = \mathbb{C}$, any skew-Hermitian form B can be written as $B = iB'$, where B' is a Hermitian form.

Technically, there are two ways we could try to generalize the usual real inner product to the complex case. We could (as we have) convert it to a sesquilinear product or we could just keep the same definition. However, it turns out that there are no positive-definite bilinear forms on a complex vector space. The sesquilinear generalization is therefore the natural one.

Let B be a positive-definite bilinear form on complex space V. Then $B(cv, cv) = c^2 B(v, v)$. Since B is positive-definite, both $B(v, v)$ and $B(cv, cv)$ are real and positive. However, this can't be the case if c^2 is not real and positive. For example, pick $c = i$.

A choice of basis induces a preferred isomorphism $V \to V^*$, as does any nondegenerate, symmetric bilinear form (including a real inner product). However, a nondegenerate sesquilinear form (and, in particular, a Hermitian inner product) does *not* define a natural isomorphism between V and V^*.

As mentioned, given a basis $\{e_i\}$ of V, we can define a basis $\{e_i^*\}$ of V via $e_i^*(e_j) = \delta_{ij}$. In the real case, this is the same as the isomorphism induced by any real inner product relative to which $\{e_i\}$ is orthonormal.

A bilinear form $B : V \times V \to K$ induces two maps $f, g : V \to V^*$, depending which argument we fix. Specifically, $f(v)(w) = B(v, w)$ and $g(v)(w) = B(w, v)$. For a symmetric bilinear form, these are the same, and for a nondegenerate bilinear form they are isomorphisms. We therefore get a single (and thus preferred) isomorphism when we have a real inner product.

For a sesquilinear form, this fails. The reason is that $B(v, w)$ is conjugate-linear rather than linear in one of its arguments. If we define f and g as above, g is not a linear map from V to V^* and f is not a map from V to V^* since each $f(v)$ is a conjugate-linear map from V to $\mathbb{C}$, not a linear one. For sesquilinear forms, there *is* a natural isomorphism between V and the conjugate-dual space. The latter is defined as the space of all conjugate-linear maps $V \to \mathbb{C}$. It tends not to appear often, because we usually end up working with a specific $2n$-dimensional realification of V, in which case a Hermitian inner product decomposes in a natural way into a real inner product and a symplectic form. We'll have more to say about this later. Note that, while the notion of "conjugate-linear" is basis-independent, the notion of a "conjugate vector" (i.e. $\overline{v}$) is not. Suppose we write $v = \sum c_i e_i$ in complex basis $\{e_i\}$. The conjugate vector $\bar{v} = \sum \overline{c_i} e_i$. If we change basis to $\{e'\}$, where $e_i = M_{ij} e'_j$ is the (complex) basis-change matrix M, then $v = \sum_{ij} c_i M_{ij} e'_j$ and $\bar{v} = \sum \overline{c_i} M_{ij} e'_j$. However, the conjugate vector in this basis to v is $\sum_{ij} \overline{c_i}\,\overline{M_{ij}} e'_j$. For the notion of conjugate vector to be basis-independent, we need $\sum_{ij} \overline{c_i}\,\overline{M_{ij}} e'_j = \sum_{ij} \overline{c_i} M_{ij} e'_j$ for all choices of c and nonsingular M. This holds for real basis-change matrices but not for complex ones.

2.2. **Vectors and Matrices.**

Though it may seem odd to do so after already mentioning tensors and inner-products, let's take a step back and consider vectors and matrices at a basic level.

The terms "matrix" and "vector" are ambiguous in common usage. "Matrix" usually refers to an array of numbers in whatever field K we're using (i.e. $\mathbb{R}$ or $\mathbb{C}$), and "vector" often means an $n \times 1$ or $1 \times n$ matrix (the former being a "column vector" and the latter a "row vector").

However, "vector" is also commonly used to refer to an element of an abstract vector space. Less commonly (and mostly in physics), the term "matrix" is sometimes used to refer to a linear map between finite-dimensional vector spaces. In the case of a linear automorphism, this corresponds to a finite-dimensional linear operator (aka a $(1, 1)$-tensor). When used to refer to abstract elements in this fashion, vectors and matrices take the form of specific arrays of numbers in any given basis, and those arrays of numbers

transform in certain ways under basis changes.

In physics, this sloppy usage arises because we're often working in a highly constrained environment where we've (explicitly or implicitly) fixed a basis. This is similar to the reason that many treatments (especially at the introductory level) ignore the distinction between vectors and dual vectors; there is an implicit choice of basis or inner product (or both) which imposes a natural choice of isomorphism between the vector space and its dual space.

To make matters more confusing, the most common usage (and the one we'll adhere to) is mixed: matrices are numeric arrays, and vectors are elements of an abstract vector space. Unless we say otherwise, this is what we'll mean from now on.

We'll refer to $n \times 1$ arrays or $1 \times n$ arrays as "column vectors" or "row vectors" or "n-tuples" (of numbers) rather than "vectors" in the non-abstract case.

We'll sometimes refer to the "canonical spaces" $\mathbb{R}^n$ and $\mathbb{C}^n$. These can be thought of as either concrete vector spaces whose elements are the n-tuples of numbers or as abstract vector spaces with a specific basis $((1, 0, \dots), (0, 1, \dots)$, etc). In both cases, each element has a concrete n-tuple of numbers associated with it. The choice of view is immaterial. When we mention elements of a canonical space, we'll always mean the corresponding n-tuples of numbers.

We won't care how these n-tuples change under basis changes, because we don't perform such operations on the canonical spaces. If we pick an element of a concrete vector space, we have a fixed n-tuple associated with it. In one view, the basis is immaterial since we're not looking at coefficients in a given basis and in the other view the basis is fixed.

The canonical spaces also have standard inner products, given on $\mathbb{R}^n$ by the usual dot product $\sum a_j b_j$ and on $\mathbb{C}^n$ by $\sum a_j \overline{b_j}$. We'll often, but not always, assume these. When we do, we'll refer to the "standard inner product". In the preferred-basis view, the canonical basis is orthonormal/unitary relative to the standard inner product (we'll discuss what this means shortly).

The standard inner product is also known as the "dot product". We are more familiar with the real case, but the complex "dot product" is as just described: $\sum a_j \overline{b_j}$.

There is a theorem in linear algebra that all vector spaces of a given finite dimension n over a given field K are isomorphic to one another. One way of thinking about this is that any finite-dimensional V over $\mathbb{R}$ or $\mathbb{C}$ is isomorphic to the canonical $\mathbb{R}^n$ or $\mathbb{C}^n$, with $n = \dim V$. As we will see, the relevant isomorphisms are precisely the bases of V.

2.3. **Bases.**

In any given basis for vector space V, all of the objects defined earlier (i.e. inner products, sesquilinear forms, etc) take the form of a square matrix, as do all linear automorphisms of V. I.e., many distinct types of objects are represented by matrices. They differ in how their matrix representatives transform under a basis change.

If we're handed a matrix without being told the type of object it represents, it is meaningless to speak of how it transforms under basis changes. It's just an array of numbers. However, if we say that it is the representative of some object FOO, then we can ask what FOO looks like in other bases. If FOO looks like matrix M_1 in basis B_1 and matrix M_2 in basis B_2, then under a basis change, its matrix representative "transforms" from M_1 to M_2. I.e., based on the nature of the abstract object, we have rules for how its matrix representative transforms under a basis change. We'll reify this shortly.

A basis for V is a choice of n ordered linearly independent vectors $(e_1, \dots, e_n)$. Each $v \in V$ has a unique expression $v = \sum_i v_i e_i$ in terms of these, and the v_i's form an n-tuple of K values. I.e., a basis can be

thought of as a map from V to the canonical K^n. It is easy to see that this map is bijective and linear. I.e., it is a vector-space isomorphism. Moreover, it can be shown that every vector space isomorphism $V \to K^n$ corresponds to a basis and vice versa. I.e., there is a bijection between bases for V and vector-space isomorphisms from V to the canonical K^n.

Note that we are speaking of bases at this point, not basis *changes*. We cannot compose the basis maps, since each is from V to K^n, not from V to itself.

Whether it is more useful to define a basis as a set of vectors or as an *ordered* set of vectors (i.e. an n-tuple of vectors) depends on the purpose. Most physics treatments take the ordered set route, which tends to make the bookkeeping a bit easier in general and a lot easier when V is the direct sum of two smaller spaces that we need to treat differently (ex. temporal and spatial directions in the presence of a pseudo-Riemannian metric or horizontal and vertical directions in the tangent space to a fiber bundle). Formally, a "basis" is a set of linearly independent vectors, and an "ordered basis" is an ordered set (aka n-tuple in the finite-dimensional case) of them. We'll use "basis" to mean an ordered basis and take a brief detour later to discuss unordered bases.

If we change basis from $(e_1, \ldots, e_n)$ to $(f_1, \ldots, f_n)$, the corresponding isomorphism from V to K^n changes as well. Denote by B_e and B_f the isomorphisms $V \to K^n$ corresponding to bases e and f. Then $B_f \circ B_e^{-1}$ is an isomorphism from K^n to itself. It tells us how the coefficient n-tuple for any given $v \in V$ changes when we move from basis e to basis f. Given an $n \times 1$ matrix c in K^n, $B_e^{-1}(c)$ is the vector $v \in V$ that has the coefficient n-tuple c in basis e. $B_f \circ B_e^{-1}$ is just the new coefficient n-tuple c' of that same v in basis f.

By choosing to represent our coefficients as $n \times 1$ matrices rather than $1 \times n$ matrices, we're choosing to work with "column vectors" rather than "row vectors". This is a choice of convention and only affects the way that results are expressed. Under our convention, we can legitimately write expressions like $c' = Mc$ and $c'^\dagger = c^\dagger M$ (for square matrix M) and $x = c^\dagger c$. [Recall that $c^\dagger$ denotes the complex conjugate of c^T.] Conventionally, people usually consider row vectors in K^n to be the "vectors" (i.e. elements) of K^n. However, it is perfectly legitimate to do otherwise, as long as we are consistent in our use of matrices and (numeric) vectors. Relative to the conventional view of vectors in K^n, our notion of basis is actually a map from V to the conjugate-dual space of K^n, and everything else is adjusted accordingly. However, the standard inner product on K^n gives us a natural isomorphism between K^n and its conjugate-dual, so the distinction is moot. Since it is more unwieldy to typeset column vectors, we'll often write $(a, b, c, d, \ldots)^\dagger$ when an explicit expression for one is needed.

In the infinite-dimensional case, a "Hamel basis" is the generalization of the type of basis we are discussing. A set of vectors is linearly independent if no vector in the set can be expressed as a *finite* linear combination of others (in fact, the term "linear combination" is upgraded to include finiteness in its definition). Note that we have no notion of what an infinite linear combination means, except in a purely formal symbolic sense. A Hamel basis is a linearly independent set of vectors such that every $v \in V$ can be expressed as a finite linear combination of basis vectors. This linear combination can be shown to be unique. In certain spaces (ex. Hilbert spaces) on which a notion of convergence exists, we can also define a "Schauder basis", allowing vectors to be convergent series of other vectors. Hamel bases are larger or, equivalently, a space generated from a specific set of vectors is smaller if they constitute a Hamel basis rather than a Schauder basis. We won't concern ourselves with Schauder bases here. In the case of an infinite-dimensional space, there is no standard inner product (though we often can pretend there is one), and we can't speak of numeric vectors or matrices. However, the notions of linear and conjugate-linear maps still exist, as do those of dual and conjugate-dual spaces. B_e and B_f are linear maps and $B_f \circ B_e^{-1}$ is a linear automorphism of the canonical space. Matrix multiplication is replaced with composition.

It will prove inconvenient to pick a distinct symbol for the coefficients each time we work with a vector, so we'll refer to the coefficients of v as $\hat{v}$ when the relevant basis is clear. When we're working in more than one basis, we'll refer to $\hat{v}^e$ and $\hat{v}^f$ to clarify which basis the coefficients are in. $\hat{v}$ is an n-tuple of K-values, arranged as a column vector (aka $n \times 1$ matrix) in K^n.

As a linear map from K^n to itself, $B_f \circ B_e^{-1}$ is a specific $n \times n$ matrix of K-values. I.e., a basis-change corresponds to a square matrix, the i^{th} column of which consists of the coefficients of e_i in the f basis. Denoting this matrix $S^{(e,f)}$, we see that $\hat{v}^f = S^{(e,f)}\hat{v}^e$ for any given v. The set of basis change matrices is the group $GL(n, K)$. Note that the term "basis change" is synonymous with "basis change matrix".

Since the sets $\{e_i\}$ and $\{f_i\}$ are each linearly independent, $S^{(e,f)}$ is patently nonsingular. Its inverse is the basis-change matrix from f to e. I.e. $(S^{(e,f)})^{-1} = S^{(f,e)}$. It's also easy to show that $S^{(e,f)}S^{(f,g)} = S^{(e,g)}$. Finally, $S^{(e,e)} = I$. Note that our labeling $S^{(e,f)}$ is a bit deceptive. For a given e or a given f, each $S^{(e,f)}$ is distinct. However, a given $S \in GL(n, K)$ takes on the role of many $S^{(e,f)}$'s. In fact, any given $S \in GL(n, K)$ takes each e to a unique f and takes a unique g to e.

We can view a given "basis change" as an action on the space of bases (aka isomorphisms from V to the canonical K^n. This action is uniquely defined by the associated matrix. If we try to do so abstractly, we'll just end up back at the nonsingular matrices.

The set of basis changes is bijective with the set of bases but has an algebraic structure via matrix multiplication (or composition in the infinite-dimensional case). This structure makes it the group $GL(n, K)$. It is important to note that the set of bases does *not* have a natural group structure. There is no notion of composition of bases and there is no "identity" basis. However, it is a torsor of $GL(n, K)$. We'll now describe what that means.

Some treatments adopt the convention that the "basis change matrix" from e to f is $B_e \circ B_f^{-1}$ instead. The only effect this has on the development below is that we must replace S with S^{-1} everywhere.

2.3.1. *Aside: Set of bases as a torsor.*

Let $L(V)$ denote the set of bases for V. It is a K-analytic manifold and its points are the ordered bases (or, equivalently, the bijective linear maps from V to the canonical K^n). By construction, we have a natural action of the group $GL(n, K)$ on $L(V)$.

If we pick a reference basis b, then the elements of V are bijective with the nonsingular square n-matrices. We therefore can impose the same manifold structure as $GL(n, K)$. It is easy to show that different choices of b yield the same maximal atlas, and we get a single K-analytic manifold structure. As far as set theory, topology, and differential geometry go, $L(V)$ is the same as $GL(n, K)$. The difference is that $GL(n, K)$ is a group, and $L(V)$ is not. However, as we will see, for each given reference b, it *is* a copy of $GL(n, K)$ in a natural way — but this copy varies with b.

If we view $L(V)$ as a set of isomorphisms $V \to K^n$, the relevant action of $GL(n, K)$ on $L(V)$ is given by $B' = S \circ B$ (where we view matrix S as an automorphism $K^n \to K^n$). I.e. $\rho(S)(B) = S \circ B$. This action is free and transitive, making $L(V)$ a "torsor" of $GL(n, K)$ as we'll now describe.

Recall that an action of group G on manifold M is a homomorphism $\rho : G \to Diff(M)$. An action is (i) "free" iff when $g \neq e$, $\rho(g)(x) \neq x$ for every x (i.e. every nontrivial g moves every point), (ii) "transitive" iff for every $x, y \in M$, there is some $g \in G$ s.t. $\rho(g)(x) = y$ (i.e. we can move between any two elements), and (iii) "effective" if every $g \neq e$ has some x s.t. $\rho(g)(x) \neq x$ (i.e. every g moves some point). An effective action ρ is injective to $Diff(M)$, and free implies effective.

Suppose ρ is effective but not injective. Then some $\rho(g) = \rho(g')$, and $\rho(g^{-1}g') = \rho(g)^{-1}\rho(g')$ since it is a homomorphism. However, the latter is just Id_M since $\rho(g) = \rho(g')$. Therefore, $\rho(g^{-1}g')$ doesn't move any points, even though $g^{-1}g' \neq e$, making the action noneffective and violating our premise. It is obvious from the definitions that free implies effective.

A torsor is a free, transitive, and smooth action ρ of a Lie group G on a smooth manifold M. It is easy to see that ρ defines a diffeomorphism between G and M.

What we mean by "smooth" here is not that ρ is smooth to $Diff(M)$ as a manifold, the very definition of which raises a host of issues. Rather, we mean that each $\rho(g)$ is a smooth automorphism of M and that for each $x \in M$, $\rho(-)(x) : G \to M$ is a smooth map. This will suffice for our purposes, but is, in fact, overkill. Basically, we'll say that ρ is smooth in whatever sense we need it to be, without delving too deeply into the precise conditions needed. All the examples we'll care about satisfy any conditions we care to throw at them.

Pick any $x \in M$, and define $\alpha : G \to M$ via $\alpha(g) \equiv \rho(g)(x)$. We need ρ to be smooth in whatever sense will make α smooth for every x, which is why we chose the condition that we did. Suppose $\rho(g)(x) = \rho(g')(x)$ for some $g' \neq g$. Then $\rho(gg'^{-1})(x) = x$, but $gg'^{-1} \neq Id_M$. This violates the assumption of freeness, since gg'^{-1} must move x. Therefore, a free action implies injectivity of α. Since ρ is transitive, $\rho(g)(x) = y$ for every y and some g, so α is surjective. We thus have a smooth bijection. It is not hard to show that the inverse must be smooth too, and we get a diffeomorphism.

However, as is evident in our comment above, we have one such diffeomorphism for each choice of $x \in M$. If we pick a given $x \in M$, then we can use α to define an isomorphic copy of G on M. We just define $y \cdot z \equiv \alpha(\alpha^{-1}(y) \cdot \alpha^{-1}(z))$. By construction, $\alpha^{-1}(x) = e$, the identity element of G. It is often said that a

torsor is a group sans the identity element, and this is what is meant. Once we choose a point $x \in M$ to serve as the identity element, we lock down a unique copy of G on M.

As a simple example, consider the unit circle in $\mathbb{R}^2$, $\{x \in \mathbb{R}^2; |x| = 1\}$. Topologically and as a manifold, this is just S^1 and has no group structure or preferred element. $\mathbb{R}^2$ comes equipped with the standard inner product, so we expect the unit circle to have a notion of "the angle between two points". I.e., we can take a difference or "delta" between two points, even if we don't have a notion of addition. In this sense, we can talk about the delta $(x - y)$. It is $-(y - x)$, and $(x - x) = 0$, and $(x - y) + (y - z) = (x - z)$. The deltas form a group under addition, specifically $\mathbb{R}$ modulo 2π.

We also have a notion of adding a delta to a point: $z + d$ is another point, and $y + (x - y) = x$ and $(x - y) + (y - z) = (x - z)$ and $(x - y) + (y - x) = 0$ (where the latter two expressions involve only deltas). I.e. S^1 looks a lot like a group but isn't quite one.

What are the deltas? They are simply the action of a group G (which is $\mathbb{R}$ modulo 2π under addition) on S^1. Specifically, $x + d = \rho(d)(x)$ in our notation. This action is (by construction) free and transitive. For any choice of $x \in S^1$, we therefore have a diffeomorphism $\alpha : G \to S^1$, given by $\alpha(d) = x + d$. This choice of preferred point (ex. north pole), allows us to endow S^1 with the same group structure as G, by serving as the identity element. We then have that $y + z \equiv \alpha(\alpha^{-1}(y) \cdot \alpha^{-1}(z)) = x + ((y - x) + (z - x))$, where the element in parentheses is a delta.

We can turn S^1 into a group, but only by choosing a point to serve as the identity element. Other examples of torsors include the fibers of principal bundles and the bases of a vector space.

In the case of bases, our manifold M is $L(V)$, and our group G is the basis change matrix group $GL(n, K)$. As with the circle, we have a notion of a group of "deltas". Given bases B_f and B_g, the delta $B_f \circ B_g^{-1}$ is a basis change matrix. The counterpart of adding a delta to a point in the circle is applying a basis change to a basis. I.e., $S \circ B_f$ rather than $x + d$, with S now playing the role of d, and B_f now playing the role of x.

A preferred point in $L(V)$ is a choice of basis b. Given such a point, $\alpha : GL(n, K) \to L(V)$ is given by $\alpha(S) = S \circ B_b$. In terms of the basis b, B_b takes the form of the identity matrix, so $\alpha(S)$ is the basis f whose matrix representative of B_f in basis b is S. Equivalently, it is the unique f s.t. $S^{(b,f)} = S$. It follows that $\alpha^{-1}(f) = S^{(b,f)}$. We can impose the group structure on $L(V)$ via $f \cdot g \equiv \alpha(\alpha^{-1}(f) \cdot \alpha^{-1}(g)) = \alpha(S^{(b,f)} S^{(b,g)})$. I.e., it is the unique basis h whose matrix representative of B_h in basis b is $S^{(b,f)} S^{(b,g)}$.

$S^{(b,b)} = I$, so $\alpha(I) = b$. f^{-1} is the unique basis s.t. $\alpha(S^{(b,f)} S^{(b,f^{-1})}) = b$, so $S^{(b,f^{-1})} = (S^{(b,f)})^{-1} = S^{(f,b)}$.

This is what it means to say that $L(V)$ is a torsor of $GL(n, K)$. Given any choice of preferred basis (to serve as the identity element), we have a natural copy of $GL(n, K)$ on $L(V)$ — but that copy varies with the choice of preferred basis.

2.4. **Orthogonal vectors.**

The terminology surrounding orthogonality and unitarity can be wildly inconsistent and confusing. Let's begin with the real case, which is (slightly) cleaner.

2.4.1. *Real case.*

A real inner product $g(v, w)$ endows V with a notion of length, embodied in the induced norm: $|v| \equiv \sqrt{g(v, v)}$, and a notion of angle, given by (for nonzero vectors) $\cos\theta_{v,w} \equiv g(v, w)/(|v| \cdot |w|)$. A given v is a **unit vector** if $|v| = 1$. Nonzero vectors v and w are **orthogonal vectors** if $g(v, w) = 0$.

Recall that a norm || on a vector space satisfies $|v| \geq 0$ for all v (with equality iff $v = 0$), $|v + w| \leq |v| + |w|$, and $|cv| = |c| \cdot |v|$, for all $c \in K$.

Some people define the zero vector to be orthogonal to every vector. We won't do this and will confine the notion of orthogonality to nonzero vectors.

A set of vectors $\{v_i\}$ is **orthogonal** if the vectors are nonzero and pairwise orthogonal (i.e. any two elements are orthogonal vectors). We define an **orthogonal basis** as a basis which is orthogonal as a set. An **orthonormal basis** is an orthogonal basis comprised solely of unit vectors.

The notions of orthogonal vectors, orthogonal bases, and orthonormal bases all require the presence of an inner product on V and depend on the particular choice of g.

Two vectors may be orthogonal under one inner product but not another.

When it comes to matrices, there is also a notion of "orthogonality", but it is defined without reference to any inner product. This can be viewed either as implicitly assuming the standard inner product on the canonical $\mathbb{R}^n$ or as simply defining the relevant notions in terms of the properties of matrices (as arrays of values).

Two nonzero column vectors or two nonzero row vectors are **orthogonal** if their dot product is 0. I.e., $\sum v_i w_i = 0$. We say that a column vector or row vector is a **unit column vector** or **unit row vector** if $\sum v_i^2 = 1$. We say that a set of row vectors or a set of column vectors is **orthogonal** if its members are pairwise orthogonal.

We can identify two special classes of real square matrices: (i) matrices M such that both $M^T M$ and MM^T are nonsingular and diagonal, and (ii) matrices M such that $M^T M = I$. Clearly, (ii) implies (i). However, the converse does not hold. (i) is a larger class of matrices than (ii).

It is easy to see that if $M^T M = D$ is diagonal, it must be nonnegative-definite. Denoting the columns of M by $v^{(i)}$, the i^{th} diagonal element of $M^T M$ is just $\sum_j (v_j^{(i)})^2$, which is nonnegative since M is real. Moreover, it can only equal 0 if $v^{(i)} = 0$. This means that D is not just nonsingular, but positive-definite. The same argument holds for MM^T, but using the rows rather than the columns.

We needn't specify $MM^T = I$ separately in (ii), because $MM^T = I$ iff $M^T M = I$. Note, however, that MM^T is *not* the transpose of $M^T M$. $M^T M = I$ implies $M^T = M^{-1}$, but the matrix inverse is two-sided, so $MM^{-1} = I$ and $MM^T = I$. However, this reasoning does not extend to (i). MM^T need not be nonsingular and diagonal just because $M^T M$ is. It turns out that any M for which $M^T M$ is nonsingular and diagonal can be written $M = M'D$ for some M' s.t. $M'^T M' = I$ (i.e. M' satisfies (ii)) and D is diagonal and positive-definite. However, MM^T then becomes $M'D'D'^T M'^T = M'DM'^T$ with a diagonal $D = (D')^2$. This need not be diagonal. We therefore must stipulate that both $M^T M$ and MM^T are diagonal and nonsingular.

As a counterexample, let $M = \begin{pmatrix} 2 & 3 \\ 3 & -2 \end{pmatrix}$. $M^T M = MM^T = \begin{pmatrix} 13 & 0 \\ 0 & 13 \end{pmatrix}$, which is nonsingular and diagonal, but does not equal I. Therefore, M satisfies (i) but not (ii).

Since (i) has rows (or columns) which are orthogonal and (ii) has rows (or columns) which are both orthogonal *and* unit row (or column) vectors, it would make sense to call the matrices which satisfy (ii) "orthogonal" and those which satisfy (i) "orthonormal". Unfortunately, life isn't so simple. The terms

orthogonal matrix and **orthonormal matrix** refer to the same thing: $M^T M = I$. There is no standard name for the matrices of (i), so we'll call them **scaled-orthogonal** for lack of a better term.

To avoid creating confusion, we'll try to avoid the use of the term "orthogonal" for matrices, favoring "orthonormal" instead. However, there are some places where it is unavoidable if we wish to make contact with standard usage.

If having two terms for (ii) and none for (i) seems absurd, it is. No doubt, there is some historical reason for this state of affairs.

For a given n, the orthogonal/orthonormal $n \times n$ real matrices form the **orthogonal group** group under matrix multiplication, denoted either $O(n)$ or $O(n, \mathbb{R})$. This is a subgroup of $GL(n, \mathbb{R})$, but is not normal in it. The scaled-orthogonal $n \times n$ real matrices form a subset of $GL(n, \mathbb{R})$ and a superset of $O(n, \mathbb{R})$ but are not a group. That is one of the principle reasons they aren't as useful. This and a number of other results are codified in the following proposition.

Prop 2.1: Let M and N be real $n \times n$ matrices. Then:

- (i) If M is orthonormal, $M^{-1} = M^T$ is orthonormal.
- (ii) If M and N are orthonormal, MN is orthonormal.
- (iii) $O(n)$ is a group under matrix multiplication.
- (iv) If M is scaled-orthogonal, then M^T and M^{-1} are scaled-orthogonal.
- (v) If M is scaled-orthogonal, it may be written $M = M'D'$ for some orthonormal M' and nonsingular diagonal D'.
- (vi) If M is scaled-orthogonal, it may be written $M = D'M'$ for some orthonormal M' and nonsingular diagonal D'.
- (vii) If M and N are scaled-orthogonal, MN need *not* be scaled-orthogonal.

Pf: (i) If $M^T M = I$, then $M^{-1} = M^T$. Multiply by M on the left to get $I = MM^{-1} = MM^T$. Therefore, M^T is orthonormal. Since $M^T = M^{-1}$, M^{-1} is orthonormal.

Pf: (ii) $(MN)^T(MN) = N^T M^T MN = N^T IN = I$.

Pf: (iii) From (i) and (ii) we know that $O(n)$ is closed under matrix multiplication and the matrix inverse. $I^T I = I$ and $IM = MI = M$, so I is the identity and is an element as well. Associativity follows from the associativity of matrix multiplication.

Pf: (iv) Since M is scaled-orthogonal, $M^T M$ and MM^T both are nondegenerate and diagonal. This is the exact same condition as for M^T to be scaled-orthogonal. Now, consider M^{-1}. $M^{-1}(M^{-1})^T = (M^T M)^{-1}$ since $(M^T)^{-1} = (M^{-1})^T$ for matrices. I.e., $M^{-1}(M^{-1})^T$ is nonsingular and diagonal iff $(M^T M)^{-1}$ is, which is true iff $M^T M$ is. The same reasoning for MM^T tells us that $(M^{-1})^T M^{-1}$ is nonsingular and diagonal.

Pf: (v) Denote the columns of M by $v^{(i)}$. Since M is scaled-orthogonal, $M^T M = D$ for some positive-definite diagonal D. For convenience, let's use $||$ to denote the canonical norm on $\mathbb{R}^n$ (i.e. $|v| = \sqrt{\sum v_i^2}$). Clearly, the i^{th} diagonal element is $D_i = |v^{(i)}|^2$. Define M' to be M, but with every column divided by the corresponding $|v^{(i)}|$. By construction, this is orthonormal. Define D' to be diagonal with $D'_i = |v^{(i)}|$ (i.e. $D' = \sqrt{D}$ in the obvious sense). Then $M = M'D'$, with M' orthonormal and D' positive-definite and diagonal.

Pf: (vi) We apply the argument of (iv) to M^T to get $M^T = M''D''$ (since $MM^T = D$ for some D). Then, we take the transpose to get $M = D''^T M''^T$. I.e., $D' = D''^T$ and $M' = M''^T$ obtained this way. Equivalently, we replicate the argument (iv) using the rows of M. We get that M' is the orthonormal matrix obtained from M by normalizing each row, and D' has the row norms as its diagonal values.

Pf: (vii) Let M and N be scaled-orthogonal. Then $(MN)^T(MN) = N^T M^T MN = N^T DN$ for some D. However, an orthonormal transform of a diagonal matrix need not be diagonal, so $(MN)^T(MN)$ need not be diagonal. This is why the scaled-orthogonal matrices do not form a group.

Counterexample for (vii): We'll need at least 3×3 matrices to exhibit a counterexample. Let $M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & \sqrt{3} \end{pmatrix}$, which yields $M^T M = D$ with D having $(1, 2, 3)$ as its diagonal elements. Let $N = \frac{1}{\sqrt{6}} \begin{pmatrix} \sqrt{2} & 1 & \sqrt{3} \\ -\sqrt{2} & 2 & 0 \\ \sqrt{2} & 1 & -\sqrt{3} \end{pmatrix}$ (which actually is orthonormal, not just scaled-orthogonal). Then $(MN)^T(MN) = \begin{pmatrix} 2 & 0 & -(\sqrt{3} * \sqrt{2})/3 \\ 0 & 2 & -\sqrt{3}/3 \\ -(\sqrt{3} * \sqrt{2})/3 & -\sqrt{3}/3 & 2 \end{pmatrix}$, which isn't diagonal. Note that this example does not pose an obstruction to $O(n)$ being a group. N is orthonormal, but M is not and can't be made orthonormal via an overall scaling (i.e. multiplication by a scalar). Every row and column of N has the same norm, but the rows and columns of M do not. Therefore, we don't have the product of orthonormal matrices, just of an orthonormal matrix and a scaled-orthogonal one.

2.4.2. *Complex case.*

Let's now consider the appropriate complex counterparts to these real notions. As mentioned earlier, there are no positive-definite bilinear forms on a complex vector space. The only generalization of the real inner product is to a Hermitian inner product $I(v, w)$.

Given a Hermitian inner product $I(v, w)$, we have an induced norm $|v| \equiv \sqrt{I(v, v)}$, just as before. We therefore can define **unit vectors** just as in the real case. Similarly, we can say that nonzero v and w are **orthogonal vectors** if $I(v, w) = 0$.

Using this notion of orthogonal vectors, we can likewise define counterparts to the real notions of **orthogonal basis** and **orthonormal basis**. However, it turns out that only orthonormal bases are of interest in the complex case. These are more often called **unitary bases**, but both terms are in common use.

When it comes to matrices, we can define four classes: (i) $M^T M$ and MM^T are nonsingular and diagonal, (ii) $M^T M = I$ (which implies $MM^T = I$), (iii) $M^\dagger M$ and $MM^\dagger$ are nonsingular and diagonal, and (iv) $M^\dagger M = I$ (which implies $MM^\dagger = I$). The first two are exactly the same as the real notions, and the latter two modify them via complex conjugation. Once again, (i) implies (ii) but not vice versa. Similarly, (iii) implies (iv) but not vice versa.

Because there are no positive-definite bilinear forms on a complex vector space — and because (i) and (ii) embody notions of scaled-orthogonality and orthonormality that would be compatible with such a form — (i) and (ii) turn out to be of little interest for complex vector spaces. (iii) and (iv) embody notions that are compatible with a Hermitian inner product, so they are the natural classes of interest. Nonetheless, (ii) does crop up from time to time.

Matrices that satisfy (iv) are termed **unitary matrices**. As before, the matrices of (iii) have no name, and we'll call them **scaled-unitary matrices** for want of a better term.

As in the real case, it is easy to see that if $M^\dagger M = D$ is diagonal, it must be nonnegative-definite. Denoting the columns of M by $v^{(i)}$, the i^{th} diagonal element is just $\sum_j |v_j^{(i)}|^2$, which must be nonnegative. We can only have a 0 diagonal element if the corresponding $v^{(i)} = 0$. Once again, we'll find it more useful to exclude this case, which means that (a) D is nonsingular, and (b) D is positive-definite. The same argument holds for $MM^\dagger$, though in this case we have to work with the rows of M.

Along similar lines to the real case, it is easy to show that $MM^\dagger = I$ iff $M^\dagger M = I$. [In fact, when N and M are nonsingular and square, it's true that $NM = I$ iff $MN = I$, because the matrix inverse is two-sided.] As before, $MM^\dagger$ need not be nonsingular and diagonal just because $M^\dagger M$ is. It turns out that any M for which $M^\dagger M$ is nonsingular and diagonal can be written $M = M'D$ for some unitary M' and a D that is diagonal and positive-definite. However, $MM^\dagger$ then becomes $M'D'D'^\dagger M'^\dagger = M'DM'^\dagger$ for diagonal matrix $D = D'D'^\dagger$. This need not be diagonal. We therefore must stipulate that both $M^\dagger M$ and $MM^\dagger$ are diagonal and nonsingular.

The unitary $n \times n$ matrices form the **unitary group**, denoted $U(n)$. It is a subgroup of $GL(n, \mathbb{C})$, but is not normal in it. Note that $U(n)$ is a *real* Lie group, despite having complex matrices as its elements.

I.e., it can only be parametrized in a real-analytic way, not a complex-analytic one. This is hinted at by its real dimension, which is n^2 and thus not always even. $U(n)$ turns out to never be a complex Lie group, including when n is even.

The scaled-unitary matrices form a subset of $GL(n, \mathbb{C})$ and a superset of $U(n)$ but are not a group.

The argument is the same as in the real case.

To make things more confusing, people sometimes do work with class (ii), the matrices for which $M^T M = I$. These form the (complex) orthogonal group, which is denoted $O(n, \mathbb{C})$. It is a subgroup of $GL(n, \mathbb{C})$ but is not related to $U(n)$. Unlike $U(n)$, it is a complex Lie group.

The real dimension of $GL(n, \mathbb{C})$ is $2n^2$, the real dimension of $O(n, \mathbb{C})$ is $n(n-1)$, and the real dimension of $U(n)$ is n^2.

To see the difference between $U(n)$ and $O(n, \mathbb{C})$, consider $n = 2$. To start with, they have different real dimensions: 2 for $O(2, \mathbb{C})$ and 4 for $U(2)$. $GL(2, \mathbb{R})$ consists of all matrices of the form $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $a, b, c, d \in \mathbb{C}$ and $ad - bc \neq 0$. $O(2, \mathbb{C})$ consists of all matrices of the form $\begin{pmatrix} \cos z & \sin z \\ -\sin z & \cos z \end{pmatrix}$, where z is complex. [It helps to note that $\cos^2 z + \sin^2 z = 1$ for complex z as well.]. On the other hand, a unitary 2×2 matrix takes the form $\begin{pmatrix} z & z' \\ -e^{i\theta}\overline{z} & e^{i\theta}\overline{z'} \end{pmatrix}$, where $|z|^2 + |z'|^2 = 1$ and θ is real. Note that the non-holomorphism is clear from the presence of complex conjugates and the spare real parameter. If this was a holomorphic parametrization (as is the one for $O(2, \mathbb{C})$), it would involve only complex variables and no conjugates.

The complex counterpart to proposition 2.1 holds. We'll state it, but won't replicate the proof (which is materially the same).

Prop 2.2: Let M and N be complex $n \times n$ matrices. Then:

- (i) If M is unitary, $M^{-1} = M^\dagger$ is unitary
- (ii) If M and N are unitary, MN is unitary.
- (iii) $U(n)$ is a group under matrix multiplication.
- (iv) If M is scaled-unitary, then $M^\dagger$ and M^{-1} are scaled-unitary.
- (v) If M is scaled-unitary, it may be written $M = M'D'$ for some unitary M' and nonsingular diagonal D'.
- (vi) If M is scaled-unitary, it may be written $M = D'M'$ for some unitary M' and nonsingular diagonal D'.
- (vii) If M and N are scaled-unitary, MN need *not* be scaled-unitary.

Since scaled-orthogonal and scaled-unitary are not preserved under multiplication, we won't find them useful (ex. when discussing basis changes below), and we won't discuss them any further.

2.5. **Terminology summary.**

Because the use of the terms "orthogonal" and "orthonormal" and "unitary" in different contexts is a wildly inconsistent tangle, let's summarize the various definitions. We'll use $I(v, w)$ for either a real or complex inner product here.

- Orthogonal vectors: In a real *or complex* inner product space, $I(v, w) = 0$ (with $v, w \neq 0$).
- Orthonormal vectors: There's no such thing.
- Unitary vectors: There's no such thing.
- Unit vector: In a real or complex inner product space, $I(v, v) = 1$.
- Orthogonal basis: In a real inner product space, $I(b_i, b_j) = s_i \delta_{ij}$, where each $s_i \neq 0$.
- Orthonormal basis: In a real *or complex* inner product space, $I(b_i, b_j) = \delta_{ij}$.
- Unitary basis: In a complex inner product space, $I(b_i, b_j) = \delta_{ij}$. I.e. same as "orthonormal basis" when the space is complex.
- Orthogonal matrix: In the real *or complex* case, a square matrix M of K-values (with $K = \mathbb{R}, \mathbb{C}$) s.t. $M^T M = I$. Same as an orthonormal matrix in the real case.
- Orthonormal matrix: Has *different* meanings in the real and complex cases. In the real case, an orthogonal matrix $M^T M = I$. In the complex case, a unitary matrix $M^\dagger M = I$.
- Unitary matrix: A square complex matrix M s.t. $M^\dagger M = I$. Same as the complex definition of an orthonormal matrix.
- Orthogonal group $O(n, K)$: The group of $n \times n$ matrices of K-values s.t. $M^T M = I$, where $K = \mathbb{R}, \mathbb{C}$.
- Orthonormal group: There's no such thing.
- Unitary group $U(n)$: The group of complex $n \times n$ matrices s.t. $M^\dagger M = I$.

2.6. **Orthonormal and unitary bases.**

Although we cannot speak of an orthonormal or unitary basis without the additional structure of an inner product, we *can* speak of orthonormal or unitary basis changes. This is because each basis change has a corresponding matrix, and we can focus on the properties of this matrix. Given any basis B, we can determine the set of bases related to it by such changes.

Put another way, a given basis is or is not orthonormal (or unitary) relative to a given inner product, but this status differs with the choice of inner product. However, a basis change intrinsically is or is not orthonormal (or unitary).

Given a real vector space V, we can define an equivalence relation on its set of bases $L(V)$ via $B \sim B'$ iff the basis change matrix between B and B' is orthonormal.

It is easy to see that this is an equivalence relation. (reflexive) I is the basis change matrix from B to B and satisfies $I^T I = I$, so $B \sim B$. (symmetric) Suppose $B \sim B'$. Then the basis change matrix S satisfies $S^T S = I$. From proposition 2.1, we know that S^{-1} is also orthonormal, so $B' \sim B$. (transitive) Let $B \sim B'$ and $B' \sim B''$, with S and S' the two basis change matrices. Then $S'S$ is the basis change matrix from B to B''. $(S'S)^T (S'S) = S^T S'^T S' S = S^T I S = I$, so S'' is orthonormal, and $B \sim B''$.

The same situation holds for unitary basis changes in the complex case. In each case, we can partition the set of bases $L(V)$ into classes that are related by orthonormal (or unitary) basis changes. Note that there is no preferred class, because we have no preferred basis.

Because we have a notion of orthogonal matrices (i.e. $M^T M = I$ rather than $M^\dagger M = I$) in the complex case as well, we could define another equivalence relation $\sim'$ on $L(V)$ via $B \sim' B'$ iff the basis change matrix is orthogonal. This is distinct from the unitarity equivalence relation $\sim$. However, it is not particularly useful, for the same reason that the complex orthogonal group isn't: the positive-definite bilinear forms that would lend it utility do not exist on a complex vector space. Because of this (and to maintain some semblance of consistency with our other terminology), we'll use "orthonormality partition" to mean the unitarity partition $\sim$ in the complex case rather than the true but uninteresting orthogonality partition $\sim'$.

To avoid an excess of caveats, we'll just refer to orthonormal matrices, orthonormal bases, and the orthonormality partition of $L(V)$ in both the real and complex cases. Bear in mind that in the complex case these refer to unitary matrices, unitary bases, and the unitarity partition of $L(V)$. The orthonormality partition of $L(V)$ is completely independent of any additional structure. If we impose an inner product, then we have a preferred set of bases: those which are orthonormal relative to it. These bases form a single equivalence class.

Any given inner product selects a preferred equivalence class, but different inner products may select different preferred classes.

If we choose a basis $B_e : V \to K^n$ for V, then $B_e(e_i) = (0, \ldots, 1, \ldots, 0)^T$ (with the 1 in the i^{th} slot), and $\{e_1, \ldots, e_n\}$ map to the canonical basis for K^n. However, this doesn't mean that B_e is an orthonormal basis relative to any given inner product $I(v, w)$. $B_e(e_i)$ takes this form regardless of the presence of an inner product on V and, if one is present, of whether the e_i's are orthogonal unit vectors relative to that inner product. For any basis B_e, $\{B_e(e_i)\}$ is the same canonical set of column vectors in K^n — unit vectors which are orthogonal to one another under the standard inner product on K^n. Using dot-product notation for the standard inner product (and recalling that this means $\sum a_j \overline{b_j}$ in the complex case), in basis B_e we have $\hat{v} \cdot \hat{w} = I(v, w)$ for all v and w iff B_e is an orthonormal basis relative to I. Equivalently, we need I to be the pull-back of the standard inner product along B_e. However, this constitutes additional structure on V. We have a standard inner product on the canonical K^n, but choosing to adopt its pull-back along a given basis B_e as an inner product on V is additional information.

It is obvious that if $\hat{v}^e \cdot \hat{w}^e = I(v, w)$ for all v, w then B_e is orthonormal relative to I, since it tells us that $I(e_i, e_j) = \delta_{ij}$. In the other direction, let B_e be orthonormal relative to $I(v, w)$. $\hat{e_i}^e \cdot \hat{e_j}^e = \delta_{ij}$ always. Since B_e is orthonormal relative to $I(v, w)$, $I(e_i, e_j) = \delta_{ij}$, so the two are equal. This shows that $\hat{v} \cdot \hat{w} = I(v, w)$ for all v, w iff B_e is orthonormal relative to I. Next, consider the pull-back. Suppose that $I(-, -)$ is the pull-back of the dot-product along B_e. Then $I(v, w) = B_e(v) \cdot B_e(w)$, but the right side is (by definition), just $\hat{v}^e \cdot \hat{w}^e$. On the other hand, suppose $\hat{v}^e \cdot \hat{w}^e = I(v, w)$ for all v, w. Then $I(v, w)$ equals the pull-back of the dot-product for every v, w. Therefore, the two must be equal.

We'll next consider what the bilinear and sesquilinear forms defined earlier look like in a given basis and — more importantly — how this matrix transforms under basis changes. However, let's first take a brief detour into unordered bases.

2.7. **Aside: Unordered Bases.**

We've been working with the ordered-set definition of a basis, but let's see how things change if we work with unordered sets. An unordered basis can be thought of as an equivalence class of ordered bases, where a basis change between elements within a class involves a simple rearrangement of rows.

Let $UL(V)$ denote the set of unordered bases for V. I.e., $UL(V)$ is a set of sets and $L(V)$ is a set of ordered sets (aka n-tuples of vectors). There is a surjective map $\pi : L(V) \to UL(V)$ that takes each ordered set to the corresponding unordered one. Obviously, it is many-to-one. In fact, it maps $n!$ elements of $L(V)$ to each element of $UL(V)$.

Let's define two equivalence relations on $L(V)$. We'll let $\sim$ continue to denote the orthonormality equivalence relation discussed earlier. I.e. $B \sim B'$ iff the basis change matrix from B to B' is orthonormal. Define $B \sim' B'$ iff the basis change matrix involves a mere rearrangement of basis vectors. I.e., if $\pi(B) = \pi(B')$. Put another way, $\sim'$ just partitions $L(V)$ by the values of the π map.

The basis change matrix for a reordering of basis vectors is automatically orthonormal, so $\sim'$ implies $\sim$ and thus is a refinement of it.

If $B \sim' B'$, then B' is just a permutation of B. Call this permutation p, so $b'_i = b_{p(i)}$. Then $v = \sum c_i b_i = \sum c'_i b'_i = \sum c'_i b_{p(i)}$, so $c'_i = c_{p(i)}$. Let S be the basis change matrix from B to B'. Since $c' = Sc$, we have $c'_i = S_{ij} c_j$, and $c_{p(i)} = S_{ij} c_j$. This tells us that $S_{ij} = \delta_{p(i),j}$. I.e., to get S we just permute the rows of the identity matrix by p. $(S^T S)_{ik} = \sum_j (S^T)_{ij} S_{jk} = \sum_j S_{ji} S_{jk} = \sum_j \delta_{p(j),i} \delta_{p(j),k} = \delta_{i,k}$, so $S^T S = I$, and S is orthonormal. Since it is real, it is also unitary.

None of this is surprising, and it all follows directly from the relevant matrix groups. The group of all basis change matrices is $GL(n, K)$. In the real case, the orthogonal group $O(n, \mathbb{R})$ is a subgroup of it, and in the complex case, both $O(n, \mathbb{C})$ and $U(n)$ are subgroups of it.

Denoting by $P(n, K)$ the group of matrices that are rearrangements of the rows (or, equivalently, columns) of I, it is trivial to see that $P(n, K)$ is isomorphic to P_n, the permutation group of n elements. Moreover, $P(n, K)$ is a subgroup of both $O(n, \mathbb{C})$ and $U(n)$ in the complex case and of $O(n, \mathbb{R})$ in the real case.

Note that none of these subgroups are normal. Neither $O(n, K)$ nor $U(n)$ are normal in $GL(n, K)$, and $P(n, K)$ is not normal in either of these or in $GL(n, K)$.

The constraint that a member of $P(n, K)$ relate B and B' is stricter than the constraint that a member of $U(n)$ relate B and B'. Hence the refinement of the induced partitions. From a group standpoint, all of these groups are acting on $L(V)$.

As discussed earlier, the basis change action of $GL(n, K)$ on $L(V)$ is free and transitive.

Technically, the action of $GL(n, K)$ on $L(V)$ is defined as follows. Let $Aut(L(V))$ denote the automorphism group of $L(V)$ (i.e. all bijective maps $L(V) \to L(V)$, as a group under composition). An action of $GL(n, K)$ on $L(V)$ is a homomorphism $\rho : GL(n, K) \to Aut(L(V))$. Since the basis change matrix from B to B' is given by $B' \circ B^{-1}$, our specific action is given by $\rho_M(B) = M \circ B$ (where we view M as a linear operator on K^n) This makes sense, because M is a map from K^n to K^n and B is a map $V \to K^n$, so the result is also a map $V \to K^n$.

As mentioned, this reflects the fact that $L(V)$ is a torsor of $GL(n, K)$.

To avoid an excess of caveats, consider $K = \mathbb{C}$ (the discussion is the same for $K = \mathbb{R}$). The basis change action of $U(n)$ on $L(V)$ is free but not transitive. Otherwise, we would just have a single orthonormality class of bases. However, the action of $U(n)$ is transitive on each orthonormality class of bases. Similarly, the permutation action of $P(n, \mathbb{C})$ on $L(V)$ is free on $L(V)$ but only transitive on each permutation class of bases.

Each unitarity class is a torsor of $U(n)$, and each permutation class of bases is a torsor of $P(n, \mathbb{C})$. The quotient spaces $GL(n, \mathbb{C})/U(n)$ and $GL(n, \mathbb{C})/P(n, \mathbb{C})$ define corresponding fiber bundles, which the actions of $U(n)$ and $P(n, \mathbb{C})$ turn into principal bundles. Bear in mind, that these two quotient spaces are not groups, because the relevant subgroups are not normal in $GL(n, \mathbb{C})$.

Each element of $GL(n,\mathbb{C})/U(n)$ is a coset, and ditto for $GL(n,\mathbb{C})/P(n,\mathbb{C})$. The fact that these quotients aren't groups means that we cannot view the action of an element of $GL(n,\mathbb{C})$ as the action of an element of $U(n)$ or $P(n,\mathbb{C})$ followed by the action of an element of the respective quotient group.

If $P(n,\mathbb{C})$ was normal in $GL(n,\mathbb{C})$ *and* the relevant sequence right-split (i.e. we had a semidirect product), then there would be a suitable copy of the quotient group in $GL(n,\mathbb{C})$ and we could view the action as $M = QP$, where P is an element of $P(n,\mathbb{C})$ and Q is an element of the copy in $GL(n,\mathbb{C})$ of the quotient group. Otherwise, we just would need to consider the action of $P(n,\mathbb{C})$ followed by an element of the action of some other group $GL(n,\mathbb{C})/P(n,\mathbb{C})$. As it is, we can't even do that.

Consider a given coset $C_M \equiv \{MP; P \in P(n,\mathbb{C})\}$. The various choices of P just rearrange elements within a permutation class of bases — and they do so in a bijective fashion. Unfortunately, C_M need not take a permutation class of bases to a permutation class of bases. I.e., it cannot be viewed as a map between permutation classes of bases.

Let $[B]'$ denote a permutation class of bases, and define $C_M([B])' \equiv \{M \circ P \circ B'; P \in P(n,\mathbb{C}), B' \in [B]'\}$. However, $\{P \circ B'; P \in P(n,\mathbb{C}), B' \in [B]'\} = [B]'$ since P acts freely and transitively on $[B]'$. Therefore, we just get $C_M([B])' = \{M \circ B'; B' \in [B]'\}$. There is no reason to expect an arbitrary basis-change M to respect permutation class. It can map different elements of the same class to distinct classes. Moreover, we can't even define a multiplication between cosets because we don't have a quotient group.

However, the inability to produce a quotient group isn't an obstruction to using unordered bases. We can't speak of a basis-change matrix-class between permutation classes of bases, but we don't need one. The only thing we really need is that $\sim$ and $\sim'$ play well together, so that we *can* speak of the unitarity class of $[B]'$. Fortunately, they do. As we saw, $\sim'$ is a refinement of $\sim$.

The gist is that unordered bases would work fine, mutatis mutandis, for everything in these notes. Nonetheless, we'll stick with ordered bases. In practice, they are much easier to work with.

2.8. **How various objects transform under a basis change.**

In any given basis e, we have a map $B_e : V \to K$, and an induced dual basis e^* on V^* (defined by $e_i^*(e_j) = \delta_{ij}$) with associated map $B_{e^*} : V^* \to K$. In the presence of a specific isomorphism between V and V^*, it is common to denote by v^* the dual to v under that isomorphism.

As mentioned earlier, any e induces such an isomorphism via $e_i \to e_i^*$. The full isomorphism takes the form $\sum v_i e_i \to \sum v_i e_i^*$, so $v^*(v) = \sum_i v_i e_i^*(\sum_j v_j e_j) = \sum v_i^2$. For $K = \mathbb{R}$, this is just the norm-squared of V under the standard inner product on the canonical $\mathbb{R}^n$.

In the complex case, we may be tempted to think that e induces a 2nd isomorphism to the dual as well, via $\alpha : \sum v_i e_i \to \sum \overline{v_i} e_i^*$. This is a real isomorphism of the corresponding $\mathbb{R}^{2n}$ but is not a complex isomorphism, since it is not complex-linear. $\alpha(v + cw) = \alpha(v) + \overline{c}\alpha(v)$, making it conjugate-linear. α is an isomorphism to the conjugate-dual space rather than the dual space. We'll expand on this shortly.

A real inner product $g(v,w)$ induces a specific isomorphism between V and V^*, given by $v^*(w) = g(v,w)$ (or, since g is symmetric, equivalently by $v^*(w) = g(w,v)$), and any basis that is orthonormal relative to

g induces the same isomorphism. In such a basis, $v^*(v) = |v|^2$ (i.e. $v^*(v) = g(v,v)$).

As mentioned earlier, a Hermitian inner product $I(v,w)$ does not induce such an isomorphism. Rather, it induces an isomorphism to the conjugate-dual space. The latter is defined as the vector space of conjugate-linear maps from V to K (i.e. maps $f : V \to K$, s.t. $f(v + cw) = f(v) + \bar{c}f(w)$). Like the dual space, the conjugate-dual space is always isomorphic to V but not in a natural way. $I(-,-)$ induces a specific isomorphism. If we define $v'(w) = I(v,w)$ (or, since $I(-,-)$ is Hermitian, equivalently via $v'(w) = \overline{I(w,v)}$), v' is a conjugate-linear map, and hence a member of the conjugate-dual space. We then have $v'(v) = |v|^2$ as in the real case. As described above, any basis e induces a basis on the dual space via $e_i^*(e_j) = \delta_{ij}$ and a corresponding isomorphism to the dual, given by $e_i \to e_i^*$ and linearly extended to all of V. It also induces a basis e' on the conjugate-dual space, given by $e'_i(e_j) = \delta_{ij}$ and a corresponding isomorphism to the conjugate-dual space, given by $e_i \to e'_i$ and *conjugate-linearly* extended to all of V. I.e., $\sum v_i e_i \to \sum \overline{v_i} e'_i$. Any basis that is orthonormal relative to $I(-,-)$ induces the same isomorphism to the conjugate-dual space as does $I(-,-)$. We won't delve any further into the conjugate-dual space here.

A $(1,1)$-tensor has the form $B : V^* \times V \to K$, a $(0,2)$-tensor has the form $B : V \times V \to K$, a $(2,0)$-tensor has the form $B : V^* \times V^* \to K$, and a sesquilinear form is $B : V \times V \to K$ (but is conjugate-linear in the 2nd term rather than linear). In any basis, all four can be represented by $n \times n$ matrices of K-values. If we choose a different basis, then each is represented by another matrix. I.e., those matrices transform according to some sort of rules under a basis change.

The rule for each of the four is dictated by the invariance of $B(-,-)$ under basis changes (where the arguments are either vectors or dual vectors as appropriate), and it differs for each. I.e., even though the sesquilinear form and all three types of tensor have matrix representatives in a given basis, they differ in the way that their matrix representatives transform under a basis change.

Note that if we are handed an $n \times n$ matrix, it makes no sense to ask how it transforms under a basis change. Only if we know that it represents some object FOO, is it meaningful to ask how it transforms under a basis change. What we are really asking is how it relates to the matrix representative of FOO in some other basis. That relationship depends on what type of object FOO is.

For example, given a matrix M and a basis e, M could represent a bilinear form or a $(2,0)$-tensor or a linear operator in that basis. However, its counterpart M' in some other basis f wouldn't be the same for all three types of object.

Let M be the relevant matrix representative for one of our objects B in basis e, let $v = \sum v_i e_i$ and $w = \sum w_i e_i$ (with $v^* = \sum v_i e_i^*$ and $w^* = \sum w_i e_i^*$), and — just for the moment — let $\hat{v}$ and $\hat{w}$ denote the coefficients for whatever type of objects v and w are (i.e. $\hat{v}$ can serve as either $\hat{v}^e$ or $\hat{v^{*^{e^*}}}$, which are the same n-tuple but transform differently under basis-changes). All three of the tensor forms can be written succinctly as $\hat{w}^T M \hat{v}$, and the sesquilinear form can be written $\hat{w}^\dagger M \hat{v}$. However, this similarity is deceptive. We may be used to thinking of column vectors as dual vectors and row vectors as vectors, but that is not the case here.

When B is a $(1,1)$-tensor, $\hat{v}$ represents a dual vector and $\hat{w}$ represents a vector. When B is a $(2,0)$-tensor, both $\hat{v}$ and $\hat{w}$ represent dual vectors. When B is a $(0,2)$-tensor or sesquilinear form, both $\hat{v}$ and $\hat{w}$ represent vectors. The way that $\hat{v}$ and $\hat{w}$ transform under a basis change reflects whether each represents a dual vector or a vector.

It is a matter of convention whether we write our expressions as (taking a bit of liberty with the $\hat{v}$ notation) $v^T M w$ or $w^T M v$. The corresponding matrix representative M under one convention is the transpose of that under the other, and the relevant transformation law for M (to be described shortly) for the type of object in question would need to be adjusted accordingly. For symmetric forms B, the two expressions are identical, and no adjustment is needed. Similar considerations apply to $v^\dagger M w$ vs $w^\dagger M v$ for a sesquilinear form, but the situation is complicated by the asymmetric roles of the two arguments. We chose to place w on the left of M and v on the right in our expression as our ordering convention, and we chose the 2nd argument to be conjugate-linear in the definition of a sesquilinear form. These two choices are consistent with $B(v,w) = w^\dagger M v$, because the expression is conjugate-linear in w and linear in v. If we had chosen to place v on the left and w on the right instead, we would get $B(v,w) = v^T M \overline{w}$, because we still need to be conjugate-linear in w rather than v. If, instead, we had kept the w-on-the-left convention but chosen the first argument to be conjugate-linear in the definition of a sesquilinear form, we would get $B(v,w) = w^T M \overline{v}$, because we now must be conjugate-linear in v rather than w. If we reversed both conventions, choosing to place v on the left and also requiring a sesquilinear form to be conjugate-linear in the first argument, we get $B(v,w) = v^\dagger M w$. The first and last alternatives (i.e. $w^\dagger M v$ and $v^\dagger M w$) are those we are familiar with. The other two are equally valid under the specified choices of convention but are less elegant to work with. If we change conventions, the relevant M and its transformation law (to be described shortly) must be modified accordingly. For example, the M for the last of the four choices is the Hermitian conjugate of the M for the first, and it transforms in a manner consistent with this. If B is Hermitian, then $w^\dagger M v = v^\dagger M w$, and $w^T M \overline{v} = v^T M \overline{w}$, so we end up with only two distinct possible conventions.

The following proposition describes a number of results mentioned so far, as well as the manner in which all the objects of interest to us transform. As usual, we'll use $\hat{v}^e$ to denote the coefficients of v in basis e (as a column vector). This is the same as $\hat{v^*}^{e^*}$, so we won't distinguish the two. Our bases will be e and f, and we'll denote by S the basis-change matrix $B_f \circ B_e^{-1}$ (what we called $S^{(e,f)}$ earlier). We'll use M^e to denote the matrix representative in basis e of whatever form B we are considering. As usual, we'll refer to both real-orthonormal and complex-unitary as "orthonormal". In a few places, it will be more convenient to refer to $M^\dagger$ for both the real and complex cases (rather than distinguishing M^T in the real case and $M^\dagger$ in the complex case). Bear in mind that S is orthonormal iff e and f are in the same orthonormality class, and e is orthonormal iff we have an inner product and e is in the preferred orthonormality class relative to that inner product.

Note that there is no meaning to asking how S itself changes under a basis change. Its matrix form is basis-independent, and a given S takes every basis to some other basis. All that we can speak of is what S^2 (or SS' if they are different basis changes) looks like for a composition of basis changes. This is just the group structure of $GL(n,K)$.

Prop 2.3: Given a basis change matrix $S = B_f \circ B_e^{-1}$ from basis $\{e_1, \ldots, e_n\}$ to basis $\{f_1, \ldots, f_n\}$:

- (i) S is nonsingular.
- (ii) S is orthonormal iff e and f are in the same orthonormality class.
- (iii) Given an inner product $I(-,-)$ and any basis e in its preferred orthonormality class, the matrix representative of $I(-,-)$ in e is the identity matrix.
- (iv) $e_i = \sum S_{ji} f_j$ for each i. I.e., $\hat{e_i}^f$ is just the i^{th} column of S.
- (v) $f_i = \sum (S^{-1})_{ji} e_j$. I.e., $\hat{f_i}^e$ is just the i^{th} column of S^{-1}. If e and f are in the same orthonormality class, then $S^{-1} = S^\dagger$, and $\hat{f_i}^e$ is the complex-conjugate of the i^{th} row of S. If also $K = \mathbb{R}$, then $\hat{f_i}^e$ is just the i^{th} row of S.
- (vi) $e_i^*(f_j) = (S^{-1})_{ij}$.
- (vii) $f_i^*(e_j) = S_{ij}$.
- (viii) $e_i^* = \sum (S^{-1})_{ij} f_j^*$. I.e., $\hat{e_i^*}^{f^*}$ is just the i^{th} row of S^{-1}. If e and f are in the same orthonormality class, then so are e^* and f^* (see (x)), and $\hat{e_i^*}^{f^*}$ is the complex conjugate of the i^{th} column of S.

- (ix) $f_i^* = \sum S_{ij} e_j^*$ for each i. I.e., $\hat{f_i^*}^{e^*}$ is just the i^{th} row of S.
- (x) e and f are in the same orthonormality class iff e^* and f^* are.
- (xi) $\hat{v}^f = S\hat{v}^e$.
- (xii) $\hat{v^*}^{f^*} = (S^{-1})^T \hat{v^*}^{e^*}$. If $K = \mathbb{R}$ and e and f are in the same orthonormality class, then $\hat{v^*}^{f^*} = S\hat{v^*}^{e^*}$.
- (xiii) For $(1,1)$-tensor B (aka linear operator on V, map $V \to V$, and map $V^* \to V^*$), $M^f = (S^T)^{-1} M^e S^T$. If $K = \mathbb{R}$ and e and f are in the same orthonormality class, then $M^f = SM^eS^T$.
- (xiv) For $(0,2)$-tensor B (aka bilinear form on V, map $V \to V^*$ in two ways), $M^f = (S^T)^{-1} M^e S^{-1}$. If $K = \mathbb{R}$ and e and f are in the same orthonormality class, then $M^f = SM^eS^T$.
- (xv) For $(2,0)$-tensor B (aka bilinear form on V^*, map $V^* \to V$ in two ways), $M^{f^*} = SM^{e^*}S^T$.
- (xvi) For sesquilinear form B, $M^f = (S^\dagger)^{-1} M^e S^{-1}$. If e and f are in the same orthonormality class, then $M^f = SM^eS^\dagger$.
- (xvii) In (xiv) and (xv), if the $(0,2)$-tensor or $(2,0)$-tensor B is symmetric, antisymmetric, nondegenerate, or (for $K = \mathbb{R}$) positive-definite, then so is its representative matrix M^e in any basis e.
- (xviii) In (xvi), if the sesquilinear form B is Hermitian, anti-Hermitian, nondegenerate, or positive-definite, then so is its representative matrix M^e in any basis e.
- (xix) Given a linear operator $L : V \to W$ between two vector spaces with $n \equiv \dim V$ and $m \equiv \dim W$, let e and e' be bases for V, let f and f' be bases for W, let S_V be the $n \times n$ basis change matrix from e to e', let S_W be the $m \times m$ basis change matrix from f to f', let M denote the matrix representative of L in bases e and f (i.e. $\widehat{L(v)}^f = M\hat{v}^e$), and let M' denote the same for e' and f'. Then $M' = S_W M S_V^{-1}$.

Note that for $W = V$, (xix) is reconciled with (xiii) as follows. Given linear operator $L : V \to V$, the associated $(1,1)$-tensor $B(-,-)$ is given by $B(v^*, w) = v^*(L(w))$. Fix our basis e. We'll use M for the matrix representative of L in basis e (as in (xix)) and denote by Q the matrix representative of B in basis e (playing the role of M from (xiii)). I.e., $B(v^*, w) = \hat{w}^T Q \hat{v}$ in basis e. Then $v^*(L(w)) = (\sum_i \hat{v}_i e_i^*)((\sum_k M_{jk}\hat{w}_k)e_j) = \sum_{i,k} \hat{v}_i M_{jk} \hat{w}_k e_i^*(e_j) = \sum_{i,k} \hat{v}_i M_{jk} \hat{w}_k \delta_{ij} = \sum_k \hat{v}_j M_{jk} \hat{w}_k = \hat{v}^T M \hat{w}$. I.e., $Q = M^T$, and we expect them to transform as transposes. This is precisely what happens. Under a basis change, $M' = SMS^{-1}$ from (xix) and $Q' = (S^T)^{-1} Q S^T$ from (xiii). The transpose of the latter is $Q'^T = SQ^T S^{-1}$, and the two therefore agree.

A particular case of a $(0,2)$-tensor in (xiv) is the real inner product. When $K = \mathbb{R}$, a real inner product $g(v,w)$ is a bilinear form with certain additional properties. As such, it has a matrix representative that transforms as $M^f = (S^T)^{-1} M^e S^{-1}$. If we confine ourselves to a given orthonormality class of bases, any S between these bases is orthonormal, so M transforms as $M^f = SM^eS^T$. In any basis in the preferred class of orthonormal bases, the matrix representative of g is $M^e = I$. When $K = \mathbb{C}$, similar considerations hold for a Hermitian inner product $I(v,w)$. It has a matrix representative that transforms as $(S^\dagger)^{-1} M^e S^{-1}$. In a given orthonormality class of bases, any intra-class basis-change S is orthonormal (aka unitary), and M transforms as $SMS^\dagger$. In any member of the preferred class of orthonormal (aka unitary) bases, the matrix representative of $I(-,-)$ is the identity matrix.

Note that when $K = \mathbb{R}$ and S is orthonormal, all three types of tensors transform as $M^f = SM^eS^T$, as does the linear operator of (xix) when $W = V$. However, this is deceptive, because they all transform differently between bases that are not in the same orthonormality class.

As alluded to earlier, if we adopt a different convention (ex. $v^\dagger M w$ instead of $w^\dagger M v$), the transformation rules change. In the tensor cases, we just take the transpose. For the $(1,1)$-tensor, we get $M^f = SM^eS^{-1}$, while for the $(0,2)$-tensor and $(2,0)$-tensor the transformation rule is unchanged. In all three cases, if e and f are in the same orthonormality class, the transformation rule is unchanged (i.e. $M' = SMS^T$). In the sesquilinear case, if we adopt the v-on-left, 2nd-argument-conjugate-linear convention, then we take the transpose to get $M^f = (S^T)^{-1} M^e \overline{S^{-1}}$. If we adopt the w-on-left, 1st-argument-conjugate-linear convention, then we take the complex-conjugate to get $M^f = (S^T)^{-1} M^e \overline{S^{-1}}$. If we adopt the v-on-left, 1st-argument-conjugate-linear convention, then we take the Hermitian conjugate to get $M^f = (S^\dagger)^{-1} M^e S^{-1}$, which is the same as before. I.e., for our four convention choices, the first and fourth transformation rules are the same and the second and third transformation rules are the same. If e and f are in the same orthonormality class, then the first and fourth rules become $M' = SMS^\dagger$, while the second and third rules become $M' = \overline{S} M S^T$ (which can also be expressed $\overline{M'} = S\overline{M}S^\dagger$).

Note that (x) holds even if $K = \mathbb{C}$, where the definition of "orthonormality" is $S^{-1} = S^\dagger$.

Our proofs of (xvii) and (xviii) are sufficient, but one can also approach this question from the standpoint of the matrix-transformation rules (i.e. (xiv), (xv), and (xvi)). Since the relevant properties, when present in B, must be present in every matrix representing B, they must be preserved under the relevant basis-change rules. It is trivial to see from the transformation rules for $(0,2)$-tensors, $(2,0)$-tensors, and sesquilinear forms that symmetry and anti-symmetry are preserved in the former two cases and Hermitianity and anti-Hermitianity are preserved in the latter case. Nondegeneracy is easy too, because $\det S \neq 0$, so the only way that any of the transformed expressions can have a zero determinant is if $\det M = 0$ to begin with. For positive-definiteness, we observe that for any nondegenerate matrix X, $\{Xv; v \in V - \{0\}\} = \{v; v \in V - \{0\}\}$. I.e., scanning over all nonzero v's is the same as scanning over all nonzero (Xv)'s. For a $(0,2)$-tensor, consider $v^T(S^T)^{-1}M(S^{-1})v$, and define $w = S^{-1}v$, to get w^TMw. Since S^{-1} is nondegenerate, scanning over all nonzero v's is the same as scanning over all nonzero w's. Since M is positive-definite, $w^TMw > 0$ for all $w \neq 0$. For a $(2,0)$-tensor, we use $w = S^Tv$. We then get $v^TSMS^Tv = w^TMw$, and the same argument applies since S^T is nondegenerate. For a sesquilinear form, we use $w = S^{-1}v$ (as with the $(0,2)$-tensor). Our expression $v^\dagger(S^\dagger)^{-1}MS^{-1}v$ then becomes $w^\dagger Mw$, and the same argument applies since S^{-1} is nondegenerate.

Pf: (i), (ii): These are automatic from the definitions.

Pf: (iii) Let $I(-,-)$ be an inner product, and let e be a basis in its preferred orthonormality class. By definition, $I(e_i, e_j) = \delta_{ij}$, since the vectors in e form an orthonormal set under $I(-,-)$.

Pf: (iv), (v): These are automatic from the definitions.

Pf: (vi,vii): $e_i^*(f_j) = e_i^*(\sum_k (S^{-1})_{kj} e_k) = \sum_k (S^{-1})_{kj}\delta_{ik} = (S^{-1})_{ij}$. By the same reasoning, $f_i^*(e_j) = f_i^*(\sum_k S_{kj} f_k) = \sum_k S_{kj}\delta_{ik} = S_{ij}$.

Pf: (viii,ix): We can write $f_i^* = \sum M_{ik} e_k^*$ for some M. Since $f_i^*(e_j) = S_{ij}$, $\sum M_{ik} e_k^*(e_j) = S_{ij}$. This means that $\sum M_{ik}\delta_{kj} = M_{ij}$ is equal to S_{ij}. I.e., $f_i^* = \sum S_{ij} e_j^*$. (viii) is just (ix) applied in the opposite direction.

Pf: (x): Suppose that e and f are in the same orthonormality class, with basis-change matrix S. I.e., $e_i = \sum S_{ji} f_i$ in (iv). From (ix), we have that $e_i^* = \sum (S^{-1})_{ij} f_j^*$. I.e., $(S^{-1})^T$ is the basis-change matrix from e^* to f^*. S is orthonormal, so $S^\dagger = S^{-1}$. We can compute $((S^{-1})^T)^{-1} = ((S^T)^{-1})^{-1} = S^T$ and $((S^{-1})^T)^\dagger = \overline{S^{-1}} = \overline{S^\dagger} = S^T$, so its inverse equals its conjugate-transpose, and the basis-change matrix from e^* to f^* is unitary as well. Going the other way, if e^* and f^* are related by a basis change matrix, which we'll write as $(S^{-1})^T$, then e and f are related by basis change matrix S. If $(S^{-1})^T$ is unitary, then $((S^{-1})^T)^{-1} = ((S^{-1})^T)^\dagger$. This reduces to $S^T = \overline{S^{-1}}$, or $S^\dagger = S^{-1}$. I.e., e and f are related by a unitary basis change matrix too.

Pf: (xi): $v = \sum_i \hat{v}_i^f f_i = \sum_i \hat{v}_i^e e_i = \sum_i \hat{v}_i^e \sum_j S_{ji} f_j$. We can swap our index names in the latter sum, to make the result more transparent. This gives $\sum_{i,j} \hat{v}_j^e S_{ij} f_i = \sum_i (\sum_j \hat{v}_j^e S_{ij}) f_i$, so $\hat{v}_i^f = \sum_j \hat{v}_j^e S_{ij}$. This is just $\hat{v}^f = S\hat{v}^e$.

Pf: (xii): We do the same thing as in (xi). $v^* = \sum \hat{v^*}_i^{f^*} f_i^* = \sum_i \hat{v^*}_i^{e^*} e_i^* = \sum_i \hat{v^*}_i^{e^*} \sum_j (S^{-1})_{ij} f_j^*$. We swap the index names for clarity, to get $\sum_{i,j} \hat{v^*}_j^{e^*} (S^{-1})_{ji} f_i^*$, so $\hat{v^*}_i^{f^*} = \sum_j \hat{v^*}_j^{e^*} (S^{-1})_{ji}$. This is just $\hat{v^*}^{f^*} = (S^{-1})^T \hat{v^*}^{e^*}$.

Pf: (xiii): Our form is $B(v^*, w)$, which equals $(\hat{w}^e)^T M^e \hat{v^*}^{e^*} = (\hat{w}^f)^T M^f \hat{v^*}^{f^*}$ in the two bases. Since $\hat{v^*}$ represents a dual vector, it transforms as $\hat{v^*}^{f^*} = (S^{-1})^T \hat{v^*}^{e^*}$. Since $\hat{w}$ represents a vector, it transforms as $\hat{w}^f = S\hat{w}^e$, so we get $(\hat{w}^e)^T M^e \hat{v^*}^{e^*} = (S\hat{w}^e)^T M^f (S^{-1})^T \hat{v^*}^{e^*}$. This must hold for all v^* and w, so $M^e = S^T M^f (S^{-1})^T$, and $M^f = (S^T)^{-1} M^e S^T$.

Pf: (xiv): We follow the same regimen as (xiii), but now both $\hat{v}$ and $\hat{w}$ transform as vectors. Our form is $B(v,w)$, which equals $(\hat{w}^e)^T M^e \hat{v}^e = (\hat{w}^f)^T M^f \hat{v}^f$ in the two bases. Since $\hat{v}$ and $\hat{w}$ both represent vectors, they transform as $\hat{v}^f = S\hat{v}^e$ and $\hat{w}^f = S\hat{w}^e$, so we get $(\hat{w}^e)^T M^e \hat{v}^e = (S\hat{w}^e)^T M^f (S\hat{v}^e)$. This must hold for all v and w, so $M^e = S^T M^f S$, and $M^f = (S^T)^{-1} M^e S^{-1}$.

Pf: (xv): We follow the same regimen as (xiv), but now both $\hat{v^*}$ and $\hat{w^*}$ transform as dual vectors. Our form is $B(v^*, w^*)$, which equals $(\hat{w^*}^{e^*})^T M^e \hat{v^*}^{e^*} = (\hat{w^*}^{f^*})^T M^f \hat{v^*}^{f^*}$ in the two bases. Since $\hat{v^*}$ and $\hat{w^*}$ both represent dual vectors, they transform as $\hat{v^*}^{f^*} = (S^{-1})^T \hat{v^*}^{e^*}$ and $\hat{w^*}^{f^*} = (S^{-1})^T \hat{w^*}^{e^*}$, so we get $(\hat{w^*}^{e^*})^T M^e \hat{v^*}^{e^*} = ((S^{-1})^T \hat{w^*}^{e^*})^T M^f ((S^{-1})^T \hat{v^*}^{e^*})$. This must hold for all v^* and w^*, so $M^e = S^{-1} M^f (S^{-1})^T$, and $M^f = S M^e S^T$.

Pf: (xvi): This is materially almost identical to (xiv), except that we replace the transpose with the Hermitian dagger. $B(v,w) = (\hat{w}^e)^\dagger M^e \hat{v}^e = (\hat{w}^f)^\dagger M^f \hat{v}^f$ in the two bases. Since $\hat{v}$ and $\hat{w}$ both represent vectors, they transform as $\hat{v}^f = S\hat{v}^e$ and $\hat{w}^f = S\hat{w}^e$, so we get $(\hat{w}^e)^\dagger M^e \hat{v}^e = (S\hat{w}^e)^\dagger M^f (S\hat{v}^e)$. This must hold for all v and w, so $M^e = S^\dagger M^f S$, and $M^f = (S^\dagger)^{-1} M^e S^{-1}$.

Pf: (xvii) (a) Symmetry: $w^TMv = v^TMw$ for all w and v iff $M = M^T$. (b) Antisymmetry: $w^TMw = -v^TMw$ for all w and v iff $M = -M^T$. (c) Nondegeneracy: Suppose B is nondegenerate but M is not. Then M has a 0 eigenvalue, and $Mv = 0$ for some $v \neq 0$. In that case $w^TMv = 0$ for all w and a nonzero v, so $B(v,w)$ is not nondegenerate, violating our assumption. (d) Positive-definite: Suppose $K = \mathbb{R}$ and $B(v,v) > 0$ for all v but M has a non-positive eigenvalue λ for some eigenvector v. $B(v,v) = v^TMv = v^T\lambda v = \lambda \sum v_i^2$. Since $K = \mathbb{R}$, $\sum v_i^2 \geq 0$, so $\lambda \sum v_i^2 \leq 0$. Therefore $B(v,v) \leq 0$, violating our assumption. Note that if $K \neq \mathbb{R}$, then $\sum v_i^2$ may be negative or complex, and the argument fails.

Pf: (xviii) Hermitianity and anti-Hermitianity follow just as in (xvii) but with $-^T$ replaced by $-^\dagger$. The argument in (xvii) for nondegeneracy is unchanged. For positive-definiteness, suppose that $B(v,v) > 0$ for all v but M has a non-positive eigenvalue λ for some eigenvector v. $B(v,v) = v^\dagger M v = v^\dagger \lambda v = \lambda \sum v_i^* v_i$. However, $\sum |v_i|^2 \geq 0$, so $\lambda \sum |v_i|^2 \leq 0$. Therefore $B(v,v) \leq 0$, violating our assumption.

Pf: (xix) As a linear operator, $L(v)$ must be independent of the basis used. Let $w = L(v)$. Then $\hat{w}^f = M\hat{v}^e$ embodies the action of L on v. After the basis change to e' in V and f' in W, v and w have coefficients $\hat{v}^{e'} = S_V \hat{v}^e$ and $\hat{w}^{f'} = S_W \hat{w}^f$. We need M to transform in a way which preserves this relationship. I.e., we need $\hat{w}^{f'} = M'\hat{v}^{e'}$. Therefore, $S_W(M\hat{v}^e) = M' S_V \hat{v}^e$. This must hold for all $\hat{v}^e$, so we have $S_W M = M' S_V$, and $M' = S_W M S_V^{-1}$.

As mentioned earlier, an alternate (and more common) convention is to define the basis change matrix in the opposite direction. In that case, $S = B_e \circ B_f^{-1}$ rather than $B_f \circ B_e^{-1}$. Under this convention, we simply swap S with S^{-1} everywhere. This makes some results prettier and some uglier.

2.9. **A note on eigenvalues and eigenvectors.**

Certain properties of the representative matrices for the four types of objects under discussion (i.e. $(1,1)$-tensors, $(0,2)$-tensors, $(2,0)$-tensors, and sesquilinear forms) — such as symmetry, antisymmetry, Hermitianity, anti-Hermitianity, nondegeneracy, and positive-definiteness — are basis-independent. Otherwise, we could not speak of them as properties of tensors or sesquilinear forms. It is easy to see that the relevant matrix transformation laws preserve those properties, as they must. In general, the same cannot be said of eigenvalues, or even the determinant, trace, or index.

Recall that, for a nondegenerate (real or complex) matrix with all real eigenvalues, the "index" is just the numbers of positive and negative eigenvalues. For a degenerate matrix with all real eigenvalues, we can infer the number of zero eigenvalues as n minus the numbers of positive and negative eigenvalues. We'll just refer to the three counts together as the "index", even though this isn't standard usage. It is not defined for matrices with complex eigenvalues. Note that some people define the index to be the number of positive eigenvalues minus the number of negative eigenvalues. In this case, we cannot infer the number of zero eigenvalues — so that would have to be included explicitly. As mentioned, we'll use the term to mean the set of all three counts.

Because all four of the objects under consideration have matrix representatives in any given basis, it may be tempting to compute eigenvalues and eigenvectors for them. We can indeed do so for any square matrix, and we'll get a set of numbers and column vectors. However, to be able to meaningfully speak of the "eigenvalues" of one of the aforementioned objects B, the eigenspectrum of its representative matrix must be independent of the choice of basis. I.e., the particular transformation rule for the representative matrices of an object of type B must preserve it. Similarly, to speak of the "eigenvectors of B", the column vectors in question must transform as vectors under basis changes (i.e. $\hat{v} \to S\hat{v}$).

We needn't be greedy, of course. If we are content to speak of the "determinant" or "trace" or "index" of B, rather than its entire eigenspectrum, we need only demand that those particular derived quantities are basis-independent, even if the eigenspectrum itself is not.

Recall that the eigenvalues of a real matrix are either real or appear as complex-conjugate pairs. As a consequence, the determinant and trace of a real matrix are always real, even if some of the eigenvalues are not. Any Hermitian complex or symmetric real matrix has only real eigenvalues, as we'll see below. However, a general real matrix can have complex eigenvalues. For example, we'll see that the eigenvalues of an antisymmetric real matrix are all pure-imaginary.

Note that 0 is considered both real and imaginary.

For a complex matrix, eigenvalues need not appear in conjugate pairs. For example, consider a diagonal matrix with an arbitrary set of complex values along the diagonal.

Under any similarity transform $M \to XMX^{-1}$ (with X nonsingular, of course), the eigenspectrum of M is preserved and the eigenvectors transform as $v \to Xv$. Of the four transformation rules, only the one for $(1,1)$-tensors (and thus for linear operators) has this form.

The idea is similar to our proofs of proposition 2.3, parts (xvii) and (xviii). Suppose that $Mv = \lambda v$ and $M' = XMX^{-1}$ (with X nonsingular). Let $w = Xv$. Then $M'w = XMv = \lambda Xv = \lambda w$. I.e., M' also has eigenvalue λ, now with associated eigenvector Xv.

For a $(1,1)$-tensor, $X = (S^T)^{-1}$, so the eigenvectors transform as $v \to (S^T)^{-1}v$. The transformation rule for a linear operator (which, as we saw, has representative matrix the transpose of that for the corresponding $(1,1)$-tensor) has $X = S$, so the eigenspectrum is preserved and $v \to Sv$.

In general, the eigenspectrum of the representative matrix for a $(0,2)$-tensor, a $(2,0)$-tensor, or a sesquilinear form is not invariant under basis changes. Nor is the determinant or trace or index. However, we can still say a few things.

We get a hint at this from the transformation laws. Since $\det S = 1/\det(S^{-1}) = \det(S^T)$, and $\det(S^\dagger) = \overline{\det S}$, we see that for a $(0,2)$-tensor $\det M' = (\det M)/(\det S)^2$, for a $(2,0)$-tensor $\det M' = (\det M)(\det S)^2$, and for a sesquilinear form $\det M' = (\det M)/|\det S|^2$. Since the only constraint on S (aside from being real or complex) is that $\det S \neq 0$, the determinant can change to almost — but not quite — anything. In the real cases, $\det S$ is real, so the sign of $\det M$ cannot change. In the case of a sesquilinear form, if $\det M$ is purely real or purely imaginary, it remains so.

For the trace, the $(1,1)$-tensor has $\text{tr}\,[(S^T)^{-1}MS^T] = \text{tr}\,[S^T(S^T)^{-1}M] = \text{tr}\,M$. For a $(0,2)$-tensor, $\text{tr}\,[(S^T)^{-1}MS^{-1}] = \text{tr}\,[(S^TS)^{-1}M] \neq \text{tr}\,M$ in general (though it is equal when S is orthogonal). For a $(2,0)$-tensor, $\text{tr}\,[SMS^T] = \text{tr}\,[(S^TS)M] \neq \text{tr}\,M$ in general (though it is equal when S is orthogonal). For a sesquilinear form $\text{tr}\,[(S^\dagger)^{-1}MS^{-1}] = \text{tr}\,[(S^\dagger S)^{-1}M] \neq \text{tr}\,M$ in general (though it is equal when S is unitary).

Prop 2.4: (i) Any Hermitian or real symmetric matrix has only real eigenvalues. (ii) Any anti-Hermitian or real antisymmetric matrix has only imaginary eigenvalues.

Pf: Real symmetric/antisymmetric are special cases of Hermitian/anti-Hermitian, so we'll focus on the latter. Suppose M is Hermitian or anti-Hermitian, and $Mv = \lambda v$ for some $v \neq 0$. Then $v^\dagger Mv = \lambda v^\dagger v$. If $\lambda = 0$, we have an eigenvalue that is both real and imaginary, and the proposition holds. Suppose $\lambda \neq 0$. Then $\lambda = (v^\dagger Mv)/(v^\dagger v)$. Both sides are numbers rather than matrices. Taking the conjugate-transpose of both, we get $\lambda^\dagger = (v^\dagger M^\dagger v)/(v^\dagger v)$. $v^\dagger v$ is real and positive, so it doesn't affect anything. $\lambda^\dagger = \overline{\lambda}$. $v^\dagger M^\dagger v = \pm v^\dagger Mv$, with $+$ if M is Hermitian and $-$ if M is anti-Hermitian. We correspondingly get $\overline{\lambda} = \pm\lambda$, which tells us that λ is real for M Hermitian (the $+$ sign) and imaginary for M anti-Hermitian (the $-$ sign).

Prop 2.5: If B is a real symmetric $(0,2)$-tensor, a real symmetric $(2,0)$-tensor, or a Hermitian sesquilinear form, it has a well-defined index.

We therefore can speak of the numbers of positive, negative, and zero eigenvalues of a symmetric bilinear form, a symmetric $(2,0)$-tensor, or a Hermitian sesquilinear form. Note that it makes no sense to speak of whether a linear operator is symmetric since it has a single argument or whether a $(1,1)$-tensor is "symmetric" since it takes two different types of arguments.

Our "proof" below relies on Sylvester's law of inertia, which we won't prove. See `https://en.wikipedia.org/wiki/Sylvester%27s_law_of_inertia` for a discussion of it.

We can leverage this to define a notion of index for an antisymmetric $(0,2)$-tensor, an antisymmetric $(2,0)$-tensor, or an anti-Hermitian sesquilinear form. Specifically, if B is the tensor or sesquilinear form in question, we compute the index of $-iB$. This is an entirely different object (it's even over a different field in the first two cases, since we've moved to $\mathbb{C}$ from $\mathbb{R}$). However, this doesn't matter. The point is that we can define an "index" for B by looking at a different object: $-iB$. The latter is anti-Hermitian, so it has real eigenvalues, which are just $-i$ times B's imaginary ones. Put another way, if we write the eigenvalues of B as $\lambda_j = ia_j$, where the a_j's are all real, then we can count the numbers of positive, negative, and zero a's — and this is basis-independent. We can call this the "index" of such a B.

Pf: For the index of B to be well-defined, the matrix representative of B must have only real eigenvalues and its index must be basis independent. The matrix representative M (in any basis e) of a symmetric $(0,2)$-tensor or a symmetric $(2,0)$ tensor is symmetric, and the matrix representative M of a Hermitian sesquilinear form is Hermitian. Proposition 2.4 tells us that all three have purely real eigenvalues, so this hurdle is cleared. Let's address the tensors first. Two symmetric real matrices M and M' are said to be "congruent" if $M' = XMX^T$ for some nonsingular matrix X. The transformation laws for both types of tensor are of this form, with $X = (S^T)^{-1}$ for the $(0,2)$-tensor and $X = S$ for the $(2,0)$-tensor. Sylvester's "law of inertia" tells us that the index is the same for congruent symmetric matrices. In our case, this means that the index of M is preserved by basis-changes for both types of objects. A similar argument holds for the sesquilinear form. Two Hermitian matrices M and M' are said to be "*-congruent" if $M' = XMX^\dagger$ for some nonsingular matrix X. The transformation rule for a sesquilinear form looks like this, with $X = (S^\dagger)^{-1}$. The complex version of Sylvester's law of inertia says that if two Hermitian matrices are "*-congruent", then they have the same index. In our case, this means that the index of M is preserved by basis changes.

For symmetric $(0,2)$-tensors, symmetric $(2,0)$-tensors, and Hermitian forms, the index is therefore basis-independent. However, even in these nice cases, the eigenvalues themselves can vary with the basis. Not even the determinant or trace are basis-independent in general. For unconstrained $(0,2)$-tensors, $(2,0)$-tensors, and sesquilinear forms, all bets are off — but only when it comes to general basis changes. The following proposition tells us that the eigenspectrum is invariant (and the eigenvectors transform quite nicely) within any given orthonormality class of bases.

I.e., any nastiness occurs when moving between orthonormality classes of bases, not within them.

Prop 2.6: If we confine ourselves to bases within a single orthonormality class, then for any of the four types of objects: (i) the eigenvalues of the representative matrix are invariant under basis-changes and (ii) the eigenvectors transform as $v \to Sv$.

It follows that the trace, determinant, and index of the representative matrix are constant (and thus well-defined) within an orthonormality class.

Pf: Consider a transformation rule of the form $M' = XMY$, with X and Y nonsingular, and let $Mv = \lambda v$ for some $v \neq 0$ (but we allow $\lambda = 0$). Define $w = Y^{-1}v$. Then $M'w = XMv = \lambda Xv$. If $w = Xv$, then we have $M'w = \lambda w$. I.e., if $X = Y^{-1}$, then the eigenspectrum of M' is the same as that of M, and the eigenvectors transform as $v \to Xv$. We saw that, in general, this is not the case except for a $(1,1)$-tensor (and a linear operator, of course). However, within an orthonormality class, it holds for all four kinds. For a $(0,2)$-tensor, $M \to (S^T)^{-1}MS^{-1} = SMS^{-1}$ when $S^T = S^{-1}$, so $X = S$ works, and the eigenvectors transform as $v \to Sv$. For a $(2,0)$-tensor, $M \to SMS^T = SMS^{-1}$ when $S^T = S^{-1}$, so $X = S$ and $v \to Sv$ works here too. For a sesquilinear form, $M \to (S^\dagger)^{-1}MS^{-1} = SMS^{-1}$ when $S^\dagger = S^{-1}$, so $X = S$ and $v \to Sv$ works yet again. We already know that it works for a $(1,1)$-tensor, but let's apply the present methodology anyway. We see that $M \to (S^T)^{-1}MS^T = SMS^{-1}$ when $S^T = S^{-1}$, so $X = S$ and $v \to Sv$.

We still can't speak of the "eigenvalues of B" for three of the four types. The eigenspectrum is invariant within each orthonormality class but can vary between classes. However, if we have a preferred class, then we can *define* the eigenspectrum of B to be that of its representative matrices in this class. We also can speak of the "determinant of B" and "trace of B" and "index of B" in that case.

An inner product selects such a preferred orthonormality class of bases. In the presence of a real inner product $g(-,-)$, we can define the eigenspectrum of any $(0,2)$-tensor or $(2,0)$-tensor to be that of its representative matrix in any basis that is orthonormal relative to g. This is well-defined, since this eigenspectrum is the same in every such basis. Similarly, in the presence of a Hermitian inner product $I(-,-)$, we can define the eigenspectrum of any sesquilinear form to be that of its representative matrix in any basis that is unitary relative to $I(-,-)$.

Note that in any preferred basis, the (real or Hermitian) inner product itself appears as the identity matrix, so its own eigenvalues and eigenvectors are trivial in such bases.

Ex. in quantum mechanics, a Hermitian inner product is part and parcel of the relevant Hilbert space. We therefore can compute the eigenspectra of the Hermitian operators which represent observables. We implicitly are working in the inner product's preferred orthonormality class of bases. Eigenvectors also transform in the obvious way between bases within the preferred orthonormality class.

2.10. **Inner Products and Isometries.**

We saw that a given inner product $I(-,-)$ on V selects a preferred orthonormality class of bases: those in which it is the identity matrix Id. The following proposition tells us that any Hermitian, positive-definite matrix serves as the representative of $I(-,-)$ in some basis.

In this and the next section, for brevity we'll often say that some tensor or sesquilinear form "is" M in some basis. Obviously, we mean that it's matrix representative in that basis is M. Also, to avoid confusion with $I(-,-)$, we'll often write Id for the identity matrix where any confusion is possible.

Prop 2.7: Given an inner product $I(-,-)$ on V: (i) In any basis, its matrix representative is Hermitian and positive-definite. (ii) The bases in which $I(-,-)$ is the identity matrix (aka the orthonormal or unitary bases) form a single orthonormality class. (iii) Given any Hermitian, positive-definite matrix M, there exists a basis in which $I(-,-)$ is M.

Although the bases in which $I(-,-)$ is Id are all related by orthonormal basis changes, those in which it is some other M need not be. For example, consider a 2-dimensional real vector space with inner product g, let $M = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$, and let g be M in basis e. As will be clear from the proof below, for g to be M in two bases related by basis change matrix S, we need $S^T M S = M$ (since the transpose and conjugate-transpose are the same for real matrices). Let $S = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then $MS = \begin{pmatrix} a & b \\ 2c & 2d \end{pmatrix}$ and $S^T MS = \begin{pmatrix} a & c \\ b & d \end{pmatrix}\begin{pmatrix} a & b \\ 2c & 2d \end{pmatrix} = \begin{pmatrix} a^2+2c^2 & ab+2cd \\ ab+2cd & b^2+2d^2 \end{pmatrix}$. For this to equal M, we need $a^2+2c^2=1$, $b^2+2d^2=2$, and $ab+2cd=0$. For S to be nonsingular, we need $ad \neq bc$. The conditions for orthonormality of S would be $a^2+c^2=1$, $b^2+d^2=1$, and $ab+cd=0$. For S to be both orthonormal and satisfy $S^T MS = M$, we need $c^2=0$ and $d^2=0$ and $cd=0$, which imply that $c=d=0$, making S singular. Therefore, there is no other basis in the orthonormality class of e for which g is M. However, this uniqueness is not a universal feature. In general, $I(-,-)$ is a given M in some (possibly empty) subset of each orthonormality class of bases. We'll have more to say about this in the next section.

To elaborate on the comment above, suppose we pick some Hermitian, positive-definite $M \neq Id$. From any basis in which $I(-,-)$ is Id, the *same* S (which would be S^{-1} of the basis-change matrix above, since we're now going from Id to M rather than the other way) takes us from each basis in which $I(-,-)$ is Id to some basis in which $I(-,-)$ is M. It establishes a bijection between those two sets of bases. This may seem to imply that the bases in which $I(-,-)$ is M are related by orthonormal basis changes, contradicting our example above. Let e and e' be bases in which $I(-,-)$ is Id and let f and f' be bases in which $I(-,-)$ is M and which are obtained from e and e' via the same (necessarily non-orthonormal) S. Let O be the orthonormal basis change that takes us from e to e'. To get from f to f' we can go $f \to e \to e' \to f'$, and the corresponding basis change matrix is SOS^{-1}. Although O is orthonormal, S is not, so SOS^{-1} need not be orthonormal.

We can think of a choice of positive-definite, Hermitian $M \neq Id$ as follows. The canonical space K^n comes equipped with a standard inner product (which manifests as the usual dot product), but we can also define other inner products on it. Each of these corresponds to a choice of Hermitian, positive-definite matrix and takes the form $y^\dagger M x$ for column vectors $x, y \in K^n$. Given an inner product $I(-,-)$ on V and a choice of basis e, we have an isomorphism $B_e : V \to K^n$. The representative matrix M^e of $I(-,-)$ in basis e is just the push-forward of $I(-,-)$ along the isomorphism B_e. I.e. the inner product $\tilde{I}(x,y)$ on K^n is given by $\tilde{I}(x,y) = I(B_e^{-1}(x), B_e^{-1}(y))$. Equivalently, $I(-,-)$ is the pull-back along B_e^{-1} of the inner product $\tilde{I}(x,y) = y^\dagger M^e x$ on K^n. I.e., $I(v,w) \equiv \tilde{I}(B_e(v), B_e(w))$. I.e., the set of inner products on V and the set of inner products on the corresponding canonical space K^n are bijective, and each choice of basis (via its associated isomorphism B_e) induces a specific bijection. In this sense, we can think of the Hermitian, positive-definite matrices as labeling the inner products on V, but where the labeling scheme depends on the choice of basis.

Pf: (i) This follows immediately from the definition of the matrix representative $I(v,w) = \hat{w}^T M \hat{v}$ in any basis.

Pf: (ii) We'll see in (iii) that there exists at least one basis in which $I(-,-)$ is Id. Suppose $I(-,-)$ is Id in two bases e and f. The transformation rule for a sesquilinear form is $(S^\dagger)^{-1} Id\, S^{-1} = Id$. I.e., $S^\dagger S = Id$. This tells us that S is orthonormal. The converse is obvious too. If S is orthonormal, it takes Id to Id. As mentioned above, this is not true for some other M. If $I(-,-)$ is M in a given basis, then the set of basis change matrices under which M is invariant is not the set of orthonormal matrices in general.

Pf: (iii) We know that in any given basis, $I(-,-)$ is a Hermitian, positive-definite matrix. Let this matrix be M in basis e. We'll construct a basis change matrix S that produces Id. Applying S to e, we then get some basis f in which $I(-,-)$ is the identity matrix Id. To construct S, we use the basis change law for a sesquilinear form: $(S^\dagger)^{-1} M S^{-1} = Id$, so $S^\dagger S = M$. Any Hermitian, positive-definite matrix M has a unique Cholesky decomposition $M = AA^\dagger$, where A is lower triangular with diagonal entries that are all real and positive. Since M is nonsingular, so is A. This gives us our basis change: $S = A^\dagger$. Since we can go from M to Id and from (via the relevant S^{-1}) Id to M', we can go from M to any other Hermitian, positive-definite matrix M'. We therefore have a basis in which $I(-,-)$ is M'. [Note that, while the Cholesky decomposition is unique, there exist other suitable choices of S which are not of the Cholesky form (i.e. upper-triangular with all real, positive values on the diagonal). We can see this as follows. $I(-,-)$ is Id in a whole orthonormality class of bases, call them $\{e\}_i$ (not to be confused with individual basis vectors). Our Cholesky $S = A^\dagger$ takes us from a basis in which $I(-,-)$ is M to one in which it is Id, so S^{-1} does the opposite. We saw in our note above that a given basis-change matrix ($S^{-1} = (A^\dagger)^{-1}$ in our case) assigns to every basis $\{e\}_i$ in which $I(-,-)$ is Id a basis $\{f\}_i$ in which it is M. However, we also saw that the bases $\{f\}_i$ are not related by orthonormal basis changes in general (this was the SOS^{-1} argument in our note above). Consider $\{e\}_1$. S^{-1} takes this to some basis $\{f\}_1$. However, there is some other basis change that takes $\{e\}_1$ to each $\{f\}_i$ with $i > 1$. This basis change is not S^{-1}, but takes $I(-,-)$ from a basis in which it is Id to one in which it is M. We therefore see that there are many non-Cholesky basis change matrices which work too.]

The linear operators from V to V (aka $(1,1)$-tensors) that preserve $I(-,-)$ are termed **isometries of** $I(-,-)$. They are defined by the requirement that $I(L(v), L(w)) = I(v,w)$, where L is the linear operator in question.

The term "isometry" can be applied more broadly to nonlinear maps as well. We'll use it for linear maps here.

In a given basis e, $I(-,-)$ is some matrix M, and L is some matrix Q. From the matrix standpoint, we need $(Q\hat{w})^\dagger M(Q\hat{v}) = \hat{w}^\dagger M\hat{v}$ for all $v, w \in V$. This requires that Q be nonsingular (M already is) and that $Q^\dagger MQ = M$, which looks like the condition for the matrix representative of $I(-,-)$ to be invariant under a basis change Q^{-1} (or, equivalently, Q, since invariance in one direction implies it in the other). This is no accident. Basis-changes and linear transforms can be viewed as the passive and active forms of the action of $GL(n,K)$ on V.

Suppose we start with M and Q in basis e and then consider them in basis f, related by basis change matrix S. Our equation in basis e is $Q^\dagger MQ = M$. Applying the transform laws for sesquilinear forms and linear operators, we get $(SQS^{-1})^\dagger((S^\dagger)^{-1}MS^{-1})(SQS^{-1}) = (S^\dagger)^{-1}Q^\dagger S^\dagger (S^\dagger)^{-1}MS^{-1}SQS^{-1} = (S^\dagger)^{-1}Q^\dagger MQS^{-1}$. Since $Q^\dagger MQ = M$, we get $(S^\dagger)^{-1}MS^{-1}$, which is precisely the transformation law for M. I.e., $Q'^\dagger M'Q' = M'$. As expected, the relationship is preserved by basis changes.

Under L, and in a given basis e, a column vector $\hat{v}^e$ goes to $Q\hat{v}^e$. This is a transformation of $v \to L(v)$. On the other hand, if we look at the same v in the basis f reached from e by $S = Q$, then $\hat{v}^f = Q\hat{v}^e$. I.e., a column vector for v in f looks like the column vector for $L(v)$ in e.

If we had used the more common $f \to e$ (rather than our $e \to f$) convention when defining the basis change matrix, we would get $S^{-1} = Q$ instead. This is why people often say that a basis change looks like the inverse of an active transform.

2.11. **Symplectic Forms and Symplectic Maps.**

As mentioned earlier, a symplectic form on V is just a nondegenerate antisymmetric bilinear form. As such, its representative matrix transforms like any other $(0,2)$-tensor under a basis change. I.e., when we change basis from e to f, $M^f = (S^T)^{-1}M^eS^{-1}$.

A vector space V over field K admits symplectic forms iff it is even-dimensional (where we mean the K-dimension, a distinction which will matter to us later). A **symplectic vector space** (or **symplectic**

space), (V, ω), is a vector space with a specific symplectic form on it. A **symplectic map** is a symplectic-form preserving linear map.

To see that an even-dimension is required, let M be the matrix representative of our symplectic form in some basis e. From nondegeneracy, we know that $\det M \neq 0$. $\det M^T = \det M$, but $M^T = -M$, so $\det M^T = (-1)^n \det M$, where $n = \dim V$. Therefore, we need $(-1)^n = 1$, so n must be even. For existence, pick any basis e and let $M = J$, where $J = \begin{pmatrix} 0 & -I_n \\ I_n & 0 \end{pmatrix}$ and I_n is the $n \times n$ K-valued identity matrix. This effectively pairs the basis elements. It is nondegenerate and antisymmetric, so we just define our $(0,2)$-tensor to be the unique one which has $M = J$ in basis e.

If our spaces are (V, ω) and (W, ν), then a linear map $f : V \to W$ is a symplectic map iff $\nu(f(v), f(v')) = \omega(v, v')$ for all $v, v' \in V$. The term "symplectomorphism" is sometimes used as well, though it is typically reserved for maps between symplectic manifolds (which are to symplectic forms what Riemannian manifolds are to inner products).

In proposition 2.3, part (xix), we saw that nondegeneracy and antisymmetry are preserved under basis changes. However, we have little more than this. Since a symplectic form is antisymmetric rather than symmetric, not even the index of its matrix is preserved under general basis changes. All we know is (from proposition 2.6) that within an orthonormality class of bases the eigenvalues are invariant and the eigenvectors transform as $v \to Sv$. For inter-class basis changes, all bets are off.

Just as there is some confusing terminology surrounding orthonormality, the use of the word "symplectic" can also be misleading. Comparing a symplectic form on V to an inner product, the analogue of a (linear) isometry is a symplectic map, and the analogue of an orthogonal matrix is a symplectic matrix. This is a defect of nomenclature.

The inner product is a positive-definite Hermitian form, and its representative matrix is a positive-definite Hermitian matrix. We don't use the term "inner product matrix" or "Hermitian matrix" to refer to orthonormality. Rather, we have the distinct term "orthogonal matrix". These are the matrices which preserve Id as the matrix representative of the inner product.

This linguistic clarity does not extend to symplectic forms and matrices. A symplectic matrix is not a matrix that is nondegenerate and antisymmetric, as one would reasonably expect from its name. Rather, it is a matrix which preserves some preferred matrix representative (the counterpart of Id) of the symplectic form. I.e., the matrix representative of a symplectic form is not a symplectic matrix. A symplectic matrix is the analogue of an orthogonal matrix, not of a Hermitian matrix.

To make matters worse, unlike orthonormality, being "symplectic" is not an intrinsic property of a matrix. As we will discuss, it depends on a choice of which antisymmetric, nondegenerate matrix we regard as "special". This is the counterpart of "Id" for orthogonal matrices. Unlike inner products and orthonormality, there is no natural choice of convention for this matrix. Each choice leads to a different definition of "symplectic matrices".

Note that this choice is independent of any symplectic forms we are working with. It truly is the counterpart of Id.

Which such "special" matrix should we choose? In the case of inner products, Id is a natural choice of Hermitian, positive-definite matrix. We could pick any other Hermitian, positive-definite matrix M, consider the group of matrices which preserve it, and conclude that the group is not $U(n)$ (or $O(n)$) but is isomorphic to it. We will do precisely this, but in greater generality, below. However, there was no need for such machinations when it came to inner products. We had a ready-made and conveniently trivial choice

of M to work with. This is not the case for symplectic forms.

Note that we are not talking about a choice of inner product, just a choice of favored representative matrix. Given such an M, any $I(-,-)$ has some set of bases in which it is M. Those bases are "preferred" and the basis-changes between them are "preferred". When $M = Id$, the latter are the orthogonal matrices and the former are the orthonormal bases relative to $I(-,-)$. If we choose a different M, then we get a different set of M-preserving matrices, and a different set of bases in which $I(-,-)$ appears as M. We'll see this more clearly soon.

Although there are four common conventions (which we'll list shortly) for a symplectic counterpart to $M = Id$, any antisymmetric, nondegenerate M can serve. Because real inner products and symplectic forms are both $(0,2)$-tensors, and their matrix representatives therefore transform as such under basis changes, much of the discussion below applies to both. We'll keep things general.

We'll work with $-^\dagger$ to accommodate Hermitian inner products too. In the case of a real inner product or real symplectic form this is the same as $-^T$. A real sesquilinear form is a $(0,2)$-tensor and the sesquilinear transform law reduces to the $(0,2)$-tensor transform law when $K = \mathbb{R}$. We therefore lose nothing by allowing this generality, it simplifies notation considerably, and we avoid lots of redundant exposition and caveats. All matrices are square and of the relevant dimension.

Suppose a nondegenerate $(0,2)$-tensor or sesquilinear form B is M in basis e. In some other basis f, related by basis-change matrix S, B takes the form $M' = (S^\dagger)^{-1}MS^{-1}$. If we demand that $M' = M$ (i.e. that S preserves M), then $(S^\dagger)^{-1}MS^{-1} = M$, which means $S^\dagger MS = M$.

Prop 2.8: (i) For a given matrix M, the set of nondegenerate matrices S s.t. $S^\dagger MS = M$ forms a subgroup $G_M \subset GL(n,K)$. (ii) If M is Hermitian and positive-definite, then G_M is isomorphic to $U(n)$ when $K = \mathbb{C}$ and $O(n)$ when $K = \mathbb{R}$.

Note that we are not requiring M to be nondegenerate in part (i). Nothing in our proof of (i) requires nondegeneracy.

The reason we must distinguish between $K = \mathbb{R}$ and $K = \mathbb{C}$ in (ii) is that the set of S matrices under consideration differs. If we use $K = \mathbb{R}$, we confine ourselves to real matrices. Any real orthogonal matrix is also unitary when viewed as complex.

It may seem in the proof of (ii) below that we exhibit a *natural* isomorphism, but this isn't the case. The Cholesky decomposition that we use is indeed unique, but it is just a convenience. There are other isomorphisms (such as replacing A with $A^\dagger$, for example). Having an explicit Cholesky isomorphism is handy, but it is just one of many isomorphisms.

Pf: (i) We've been through a similar proof before, but we'll go through the motions anyway. $Id \cdot M \cdot Id = M$, so $Id \in G_M$. We already know that $S \in G_M \implies S^{-1} \in G_M$ from the expressions above (just apply $(S^\dagger)^{-1}$ to the left and S^{-1} to the right). Suppose $S, S' \in G_M$. Then $(SS')^\dagger M(SS') = S'^\dagger S^\dagger MSS' = S'^\dagger MS' = M$. Associativity follows from that of matrix multiplication.

Pf: (ii) Suppose M is Hermitian and positive-definite. As we mentioned in the proof of proposition 2.7, any positive-definite, Hermitian matrix has a unique Cholesky decomposition $M = AA^\dagger$, where A is lower triangular with real, positive diagonal entries. We also saw that A is necessarily nonsingular. Since $S^\dagger AA^\dagger S = AA^\dagger$, $(A^{-1}S^\dagger A)(A^\dagger S(A^\dagger)^{-1}) = Id$. However, this is just $N^\dagger N = Id$, with $N \equiv A^\dagger S(A^\dagger)^{-1}$. Our map $\alpha : G_m \to U(n)$ is defined to be $\alpha(S) = A^\dagger S(A^\dagger)^{-1}$. First, let's prove it's a homomorphism. $\alpha(Id) = Id$. $\alpha(S^{-1}) = (A^\dagger S^{-1}(A^\dagger)^{-1}) = (A^\dagger S(A^\dagger)^{-1})^{-1}$. $\alpha(SS') = (A^\dagger SS'(A^\dagger)^{-1}) = (A^\dagger S(A^\dagger)^{-1})(A^\dagger S'(A^\dagger)^{-1}) = \alpha(S)\alpha(S')$. Next, let's prove it's injective. Suppose $\alpha(S) = \alpha(S')$. Then $(A^\dagger S(A^\dagger)^{-1}) = (A^\dagger S'(A^\dagger)^{-1})$. Since A is nonsingular, just cancel out the right and left sides to get $S = S'$. Finally, let's prove it's surjective. Suppose $X \in U(n)$. To have $\alpha(S) = X$, we need $(A^\dagger S(A^\dagger)^{-1}) = X$. This gives us $S = (A^\dagger)^{-1}XA^\dagger$. Since all three are nonsingular, S is nonzero. The S thus obtained is obviously a member of G_M, since $S^\dagger MS = (AX^\dagger A^{-1})(AA^\dagger)((A^\dagger)^{-1}XA^\dagger) = AX^\dagger XA^\dagger = AA^\dagger = M$. A bijective homomorphism is an isomorphism. If $K = \mathbb{R}$, we restrict α to the subgroup of G_M containing real matrices, and its image is obviously $O(n) \subset U(n)$. The restricted map is an isomorphism to its image.

We've now seen that each choice of M gives rise to a subgroup of $GL(n,K)$, and that each Hermitian inner product produces a $G_M \approx U(n)$ and each real inner product produces a $G_M \approx O(n)$. These are equal (rather than merely isomorphic) to $U(n)$ or $O(n)$ iff $M = Id$.

We can say a bit more when M is nondegenerate.

Prop 2.9: Let M be nondegenerate, and let S be a basis-change (i.e. nondegenerate) matrix. (i) If $S^\dagger MS = M$, then the conditions $[M,S] = 0$ (i.e. $MS = SM$) and $S^\dagger S = Id$ are the same. (ii) If both $[M,S] = 0$ and $S^\dagger S = Id$ hold, then $S^\dagger MS = M$. (iii) The set of S's satisfying $[M,S] = 0$ is a subgroup of $GL(n,K)$. (iv) The set of S's satisfying both $[M,S] = 0$ and $S^\dagger S = Id$ is a subgroup of G_M.

I.e., if S preserves M, then either M commutes with S *and* S is orthonormal or neither holds.

Note that we don't have a full converse in (ii). It is quite possible that neither $[M,S] = 0$ nor $S^\dagger S = Id$ and yet $S^\dagger MS = M$. Requiring orthonormality and commutativity (potentially) only captures a subset of the S's which preserve M.

The elements of a group G which commute with some fixed $g \in G$ form a subgroup of G known as the "centralizer" of g. Our group in (iii) is the centralizer of M in $GL(n,K)$.

Pf: (i) Suppose $S^\dagger MS = M$. If S is orthonormal, then $S^\dagger = S^{-1}$, so $S^{-1}MS = M$, and $MS = SM$. On the other hand, if $[M,S] = 0$, then $MS = SM$, so $S^{-1}MS = M$. This means that $S^{-1}MS = S^\dagger MS$. Since everything is nondegenerate, we can cancel the MS factors from both sides, to get $S^{-1} = S^\dagger$.

Pf: (ii) Suppose that both $S^\dagger S = Id$ and $SM = MS$. Then $S^{-1}MS = M$ and $S^\dagger = S^{-1}$, so $S^\dagger MS = M$.

Pf: (iii) Clearly, $[Id,S] = 0$. Suppose $MS = SM$. Multiply on the right and left by S^{-1}, to get $S^{-1}M = MS^{-1}$. Therefore, S^{-1} commutes with M too. Suppose that $[M,S] = 0$ and $[M,S'] = 0$. Then $[M,SS'] = MSS' - SS'M = SMS' - SS'M = S(MS' - S'M) = 0$, so SS' commutes with M as well.

Pf: (iv) We saw in (iii) that the set of S's which commute with M is a group, and we know that $S^\dagger S = Id$ defines a group. The intersection of two subgroups is a subgroup, so their intersection is a subgroup of $GL(n,K)$. We saw in (ii) that this is a subset (and therefore subgroup) of G_M.

Counterexample to full converse in (ii), where $[M,S] \neq 0$ and $S^\dagger S \neq Id$, yet $S^\dagger MS = M$: Let $S = \begin{pmatrix} 3 & 0 \\ 0 & 1/3 \end{pmatrix}$ and let $M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Then S and M are nondegenerate and qualify for the proposition. $S^\dagger MS = M$, but $S^\dagger S = \begin{pmatrix} 9 & 0 \\ 0 & 1/9 \end{pmatrix} \neq Id$ and $[M,S] = \begin{pmatrix} 0 & -8/3 \\ 8/3 & 0 \end{pmatrix} \neq 0$.

Prop 2.10: For any square matrix M, $G_M = G_{M^\dagger} = G_{-M}$.

Pf: If $S^\dagger MS = M$, then it obviously holds for $-M$ too. Taking the conjugate-transpose, we get $S^\dagger M^\dagger S = M^\dagger$, so it holds for $M^\dagger$ too.

We can now see what is happening a little more clearly. We've already denoted by G_M the group satisfying $S^\dagger MS = M$. Denote by C_M the centralizer of M (defined by $[M,S] = 0$) and by Z_M the intersection $C_M \cap U(n)$. Proposition 2.9 tells us that Z_M is a (possibly proper) subgroup of G_M.

For $K = \mathbb{R}$, we're confining ourselves to real matrices — or, equivalently, working in $GL(n,\mathbb{R})$ rather than $GL(n,\mathbb{C})$. In any event, we get $Z_M = C_M \cap O(n)$ in that case. We won't be fussy with our notation, and our meaning is clear. We'll continue to speak of $U(n)$, but the appropriate substitution of $O(n)$ should be made when $K = \mathbb{R}$.

Any subgroup $G \subseteq GL(n,K)$ defines an equivalence relation amongst (and thus a partition on) the set of bases $L(V)$ via $e \sim f$ iff $S^{(e,f)} \in G$. If $G = GL(n,K)$, this partition has a single class. If $G = \{Id\}$, every basis is its own class. If $H \subset G$, we get a refinement, since $e \sim f$ under H implies $e \sim f$ under G.

A group defines an equivalence relation because (i) $e \sim e$ via $Id \in G$, (ii) if $e \sim f$ via $S \in G$, then $f \sim e$ via $S^{-1} \in G$, and (iii) if $e \sim f$ via $S \in G$ and $f \sim g$ via $S' \in G$, then $e \sim g$ via $S'S \in G$.

More generally, any action of a group G on a set X partitions X into orbits.

We have three distinct partitions of $L(V)$, based on C_M and G_M and $U(n)$. None of these three is a refinement of the other two in general, and they can be thought of as striping $L(V)$ in different ways. The partition based on $Z_M = U(n) \cap C_M$ is their common refinement.

We'll speak of Z_M, C_M, G_M, and $U(n)$ as partitions and refer to their classes, with the understanding that we mean the partition induced by the relevant group.

Each class of Z_M consists of those bases in a particular orthonormality class that are also related by S's which preserve M. Our proposition above says that each class of Z_M sits in a class of G_M as well. I.e., the partition for Z_M is the common refinement of all three others.

For lack of a better term, let's call the relevant choice of "special" matrix the "reference matrix", and change its designation to Ω instead of M (to avoid confusion with the representative matrices of some particular object, which we've been calling M^e, etc). When dealing with inner products, our choice of Ω must be positive-definite and Hermitian.

Bear in mind that Ω is just a particular matrix. There is no notion of it transforming in some fashion under basis changes. It is not the representative of some particular object in the abstract. However, in any given basis, some object(s) may have it as a matrix representative.

Suppose we've picked a positive-definite, Hermitian reference matrix Ω. Consider some Hermitian inner product $I(-,-)$ and some orthonormality class of bases. In general, $I(-,-)$ has a number of distinct representative matrices within that orthonormality class. It is Ω only on some subset of the class. If that subset is nonempty, then it equals a single class of Z_Ω.

Any two bases in the subset are related by a basis change matrix that is both orthonormal and preserves Ω. Proposition 2.9 tells us that it therefore also commutes with Ω.

If $\Omega = Id$, then the relevant subset of bases is empty except in the preferred orthonormality class of $I(-,-)$, where it constitutes the entire orthonormality class.

It is not hard to visualize what is happening. Each $I(-,-)$ doesn't pick a preferred orthonormality class; it picks a preferred class of the G_Ω partition for each Ω. When $\Omega = Id$, $G_\Omega = U(n)$, and the distinction is moot. For other Ω's, it is not.

There are several points to be made here:

- For a given Ω, G_Ω partitions the bases (i.e. $L(V)$) into classes, whose elements are related by basis-change matrices that are in G_Ω.
- For a given $S \in GL(n,K)$, there is a set of Ω's s.t. S preserves M. These are the Ω's for which $S \in GL_\Omega$.
- For a given basis e and any Ω, there is some class of G_Ω in which e sits.
- For a given $I(-,-)$, $L(V)$ is partitioned into classes based on the representative matrix of $I(-,-)$. These classes are indexed by the positive-definite Hermitian Ω's.
- For a given e, there is a specific bijection between positive-definite Hermitian matrices Ω and inner products. It takes each $I(-,-)$ to its representative matrix in that basis.

It can be a bit hard to visualize the relationship between these items. For any suitable Ω, $I(-,-)$ picks a preferred class of G_Ω. However, each G_Ω stripes $L(V)$, and these stripings criss-cross.

How do we get disjoint classes when we're picking one class from each of a large set of criss-crossing partitions? To understand this, let's consider a simple example. Let V be real and 2-dimensional, and let $g(-,-)$ be an inner product on it. Consider Id and $M = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$. Both are positive-definite and Hermitian, so each represents $g(-,-)$ in some (nonempty) set of matrices. $G_{Id} = O(2)$ partitions the bases into orthonormality classes. To parametrize the set of bases $L(V)$, pick some reference basis e and use $GL(2,\mathbb{R})$ to get from e to every other basis. I.e., we put elements of $GL(2,\mathbb{R})$ in one-to-one correspondence with elements of $L(V)$. e is assigned the identity matrix Id. This is what it means to be a torsor. By making a particular choice of basis to serve as the identity element, $L(V)$ acquires the group structure of $GL(2,\mathbb{R})$. The bases are now parametrized just like $GL(2,\mathbb{R})$: $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, with the constraint $ad - bc \neq 0$. Each orthonormal matrix has the form $\begin{pmatrix} cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$ for some θ. We thus eliminate one parameter. Put another way, each orthonormality class is a 3-dimensional submanifold of $L(V)$, and these submanifolds are parametrized by θ. $g(-,-)$ picks one of these submanifolds as its "preferred" manifold. Call this Y_{Id}. Similarly, the elements of G_M form a one-dimensional manifold (since it is isomorphic to $O(2)$). G_M partitions $L(V)$ into 3-dimensional submanifolds in a different way than $O(2)$ does. The preferred G_M-class picked out by $g(-,-)$ is disjoint from Y_{Id} but is not an orthonormality class. This means it crosses a bunch of orthonormality classes — but not Y_{Id} (since the representative matrix of $g(-,-)$ can't be both Id and M in a given basis). Call this class (the counterpart of Y_{id}) Y_M. The set of Hermitian positive-definite matrices M is a 3-dimensional manifold. Each element of it defines a G_M and associated partition of $L(V)$. There is some class Y_M of that partition which $g(-,-)$ picks out — and which is disjoint from all the other $Y_{M'}$'s (again, since g can't have more than one matrix representative in a given basis). This is a sort of tiling. For example, consider a 2x2x2 inch cube. We can view it as composed of 4 parallel 1x1x2 inch sticks in each of 3 different ways (parallel to the x-axis, parallel to the y-axis, and parallel to the z-axis). Each is a partition of the cube. We can pick a disjoint set of 3 sticks, one from each of these partitions. However, this leaves 3 1x1x1 holes, and our three removed sticks do not themselves form a partition of the cube. It may seem impossible to avoid such holes in the finite case, and it is. However, in the infinite case it is not. When we have infinitely many partitions, we can in fact stitch together these classes.

A word on counting. Consider the real case. The set of basis-change matrices has n^2 parameters. The set of symmetric, positive-definite matrices has $n(n+1)/2$ parameters. The set of orthogonal matrices has $n(n-1)/2$ parameters (just like the antisymmetric matrices — which is no accident, since one is the Lie algebra of the other). Each orthonormality class has $n(n-1)/2$ parameters, and the choice of orthonormality class is parametrized by $n(n+1)/2$ variables. The numbers therefore add up.

The same holds for any G_Ω, with Ω Hermitian and positive-definite, since this group is isomorphic to $O(n)$.

In the complex case, the set of basis change-matrices has $2n^2$ real parameters, the set of Hermitian positive-definite matrices has real dimension n^2, and the set of unitary matrices has real dimension n^2, so everything works out as well.

Why bother with all this, when we'll only ever use $\Omega = Id$ for inner products in practice? In the case of symplectic forms, we don't have a canonical choice akin to $\Omega = Id$, and we are forced to engage with G_Ω. Because symplectic forms are primarily of interest when $K = \mathbb{R}$ rather than $K = \mathbb{C}$, we'll assume that $K = \mathbb{R}$ from now on.

We can define symplectic forms over any field (albeit with some technical nuances when $charK = 2$), but a more useful concept in the complex case would be a sesquilinear form that is anti-Hermitian and nondegenerate.

Just as we must choose an Ω that is Hermitian and positive-definite when working with inner products, we must choose a reference matrix Ω that is nondegenerate and antisymmetric when working with symplectic forms. Proposition 2.8, part (i) applies to all matrices, and proposition 2.9 applies to all nondegenerate matrices, so both apply to any matrix which can represent a symplectic form. In particular, they apply to any suitable choice of reference matrix Ω for symplectic forms — the counterpart of Id. When it comes to symplectic forms, the key takeaway from these propositions is that G_Ω is a subgroup of $GL(n,\mathbb{R})$.

Given some favored choice of nondegenerate anti-symmetric matrix Ω, a **symplectic matrix** is any S that satisfies $S^T\Omega S = \Omega$. It is easy to see that such an S must be nondegenerate (and, in fact, have

$(\det S)^2 = 1$).

$\det(S^T M S) = (\det S)^2 \det M = \det M$, so $(\det S)^2 = 1$.

Note that it is defined this way for any K. I.e., for $K = \mathbb{C}$, we still use the transpose rather than the conjugate-transpose, which is why the concept is less useful in that case.

For $K = \mathbb{R}$, $S^\dagger = S^T$, so all our previous machinery carries over just fine, and the set of symplectic matrices is simply our group G_Ω. This induces a partition of $L(V)$. Each class of this partition consists of bases related by S's which preserve our special matrix Ω. Any symplectic form ω on V picks out a preferred class of bases: those in which ω is Ω. Such bases are termed **symplectic bases** or **canonical bases**. Again, their definition, depends on our choice of Ω.

Bear in mind that the set of "symplectic matrices" depends on our choice of Ω. Even apart from the confusing use of "symplectic" as the counterpart of orthogonal rather than Hermitian positive-definite, this nomenclature departs from that for inner products. "Orthogonal matrix" has a fixed meaning. We don't say that some matrix M is "orthogonal" relative to some Ω. However "symplectic" has no such fixed meaning. It requires a reference Ω. As we will see, this is often taken to be $\pm J$ (defined below) — but it need not be.

As with inner products, we can think of symplectic forms in terms of the canonical $\mathbb{R}^{2n}$. Every antisymmetric nondegenerate $2n \times 2n$ real matrix defines a symplectic form on $\mathbb{R}^{2n}$. Given V and some basis e, we have a bijection between symplectic forms on V and those on $\mathbb{R}^{2n}$ (i.e. $2n \times 2n$ real antisymmetric nondegenerate matrices) obtained via the push-forward/pull-back along the isomorphism B_e. Again, the latter are *not* "symplectic matrices", though they *are* "symplectic forms" on $\mathbb{R}^{2n}$. We therefore can think of the nondegenerate, antisymmetric $2n \times 2n$ real matrices as indexing the symplectic forms on V, though the specific choice of indexing (i.e. bijection) depends on the basis e. This is entirely analogous — unfortunate naming conventions aside — to the situation with inner products.

We write $\mathbb{R}^{2n}$, rather than $\mathbb{R}^n$, since a symplectic form requires an even dimensional space. Defining n to be half the dimension proves convenient.

As mentioned, there isn't a canonical choice of Ω when it comes to symplectic forms. However, there are four common choices.

Note that this use of the symbol J (with either sign convention) is by no means universal, and J' is purely homegrown. For example, J is also often used to designate a matrix of all 1's.

- (i) $J = \begin{pmatrix} 0 & -I_n \\ I_n & 0 \end{pmatrix}$, with I_n the $n \times n$ identity matrix.
- (ii) $-J$.
- (iii) J' is block diagonal with each diagonal block of the form $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.
- (iv) $-J'$.

In the symplectic bases, J pairs basis element e_i with e_{i+n} for $i = 1 \ldots n$, and J' pairs adjacent basis elements e_i and e_{i+1} for even i. Whether we use $\pm$ in either case is a matter of taste. The choice of J vs J' is a matter of convenience. Whether it is easier to work with a 2×2 matrix of $n \times n$ blocks or an $n \times n$ matrix of 2×2 blocks depends on the application. For some purposes, $\pm J$ is easier to use and for others $\pm J'$ is. We'll stick with J.

Many people prefer $-J$, and it is common to define J with the opposite sign convention (i.e. as $J \equiv \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}$).

We're being a little glib about the pairing. Many possible pairings lead to J and many possible pairings lead to J'. For example, pairing e_i with e_{2n+1-i} would work too. A choice of pairings of basis elements contains more information that a mere choice of representative matrix. However, the information we actually care about is contained in the representative matrix. If we care about the pairing of basis elements, we can make a choice that is convenient and consistent with the Ω in use. Usually, this is one of the two mentioned, depending on whether we prefer $\pm J$ or $\pm J'$. We won't delve into such technical ordering issues here, but we will briefly touch on them later.

Under $\Omega = J$, we get a specific group of matrices G_J which preserve it. From proposition 2.10, we know that $G_{-J} = G_J$ and $G_{-J'} = G_{J'}$. It is obvious that G_J is isomorphic to $G_{J'}$ since J and J' are related by a mere rearrangement of rows or columns. We can extend this.

The specific group $G_{\pm J}$ is called the **symplectic group** $Sp(2n, \mathbb{R})$. This is the counterpart to $U(n)$ or $O(n)$. The following is the symplectic counterpart of proposition 2.8, part (ii). It tells us that every other such G_Ω is isomorphic to $Sp(2n, \mathbb{R})$.

Prop 2.11: For any antisymmetric, nondegenerate real $2n \times 2n$ matrix Ω, G_Ω is isomorphic to $Sp(2n, \mathbb{R})$.

Obviously, not every G_Ω need be distinct. We already saw that $G_\Omega = G_{-\Omega}$, for example.

Pf: We want to construct an isomorphism between G_Ω and G_J. Suppose that $S^T \Omega S = \Omega$ (i.e. $S \in G_\Omega$) for some nonsingular S and that $\Omega = X^T J X$ for some nonsingular X. Let's first show that this would do the trick, and then we'll exhibit a suitable X. Since $S^T X^T J X S = X^T J X$, we have $(X^T)^{-1} S^T X^T J X S X^{-1} = J$. Letting $Y = (XSX^{-1})$, we have $Y^T J Y = J$, so $Y \in G_J$. We did something analogous earlier (in the case of the inner product) using the Cholesky decomposition. To show that the resulting map $S \to (XSX^{-1})$ is a group isomorphism, we first note that it is a bijection. If $(XSX^{-1}) = (XS'X^{-1})$ then we just cancel the left and right factors (which are invertible) to get $S = S'$. The map is therefore injective. For any Y, $S = X^{-1}YX$ maps to Y, so the map is surjective as well, and thus a bijection. $Id \to Id$ and $S^{-1} \to (XS^{-1}X^{-1}) = (XSX^{-1})^{-1}$, so S^{-1} maps to $(XSX^{-1})^{-1}$. $SS' \to (XSS'X^{-1}) = (XSX^{-1})(XS'X^{-1})$, so multiplication is preserved. We therefore have a homomorphism. A bijective homomorphism is an isomorphism. Now that we know this will do the trick, all that remains is to exhibit a nonsingular X such that $\Omega = X^T J X$. Since Ω is antisymmetric and nondegenerate, its eigenvalues are purely imaginary and nonzero. Nonreal eigenvalues always appear in conjugate pairs, so the eigenvalues are of the form ia_1, $-ia_1$, etc (where the a_j's are nonzero and real). Suppose we've ordered the eigenvalues as $ia_1, \ldots, ia_n, -ia_1, \ldots, -ia_n$. Recall that a matrix is "normal" if $\Omega\Omega^\dagger = \Omega^\dagger\Omega$. Hermitian, anti-Hermitian, real symmetric, and real-antisymmetric matrices are normal. Any matrix Ω can be written, via the Schur decomposition, as $\Omega = U^\dagger T U$ for some unitary matrix U and upper-triangular matrix T whose diagonal entries are the eigenvalues of Ω. We're also guaranteed (by the Schur decomposition) that if Ω is real, then U is real (i.e. orthonormal in the real sense) — though T can still be complex. There is a version of the spectral theorem for normal matrices, and it tells us that a matrix is normal iff T is diagonal. I.e., for any antisymmetric real matrix Ω, we have $\Omega = O^T D O$, where D is diagonal (with the eigenvalues of Ω as its elements), and O is real orthonormal. We can reorder the entries in D via nondegenerate, real orthonormal swap matrices. [For example, to swap the order of diagonal entries in a 2×2 diagonal matrix $D = \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}$, we use $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ (which is involutive, so $X = X^{-1}$) to get $X^{-1} D X = \begin{pmatrix} d & 0 \\ 0 & a \end{pmatrix}$.] We therefore lose no generality in writing $\Omega = O^T D O$ for any ordering of the diagonal we choose, and implicitly incorporating whatever swap matrices are needed into O. In our case, we'll use $ia_1, \ldots, ia_n, -ia_1, \ldots, -ia_n$. We can do the same for J, to get $J = V^T D' V$, where V is real orthonormal and D' consists of n copies of i followed by n copies of $-i$ along the diagonal (since these are the eigenvalues of J). Let A be defined to be the real diagonal matrix with diagonal $(\sqrt{a_1}, \sqrt{a_2}, \ldots, \sqrt{a_n}, \sqrt{a_1}, \sqrt{a_2}, \ldots, \sqrt{a_n})$. Then $D = A^T D' A$, so $\Omega = O^T A^T D' A O$. Since $J = V^T D' V$, we have $\Omega = O^T A^T (V^T)^{-1} J V^{-1} A O$. This is just $\Omega = O^T A^T V J V^T A O$, since V is orthonormal. Let $Q \equiv V^T A O$. It is nondegenerate, since V and A and O are, and $\Omega = Q^T J Q$, so we have our desired relationship.

We've now seen that any nondegenerate, antisymmetric real $2n \times 2n$ matrix Ω can serve as our reference matrix, and each defines a notion of "symplectic matrices" which yields a copy of the symplectic group in $GL(2n, \mathbb{R})$ — along with a corresponding partition of $L(V)$ into classes within which any two bases are related by a symplectic basis change matrix. This partition depends on the choice of Ω, of course.

Any given symplectic form ω picks out a preferred class in which its representative matrix is Ω. As in the case of the inner product, we can identify three subgroups, which stripe $L(V)$ in different ways. These are $Sp(2n, \mathbb{R})$, G_Ω, and C_Ω. We also can define $Z_\Omega \equiv C_\Omega \cap Sp(2n, \mathbb{R})$. Unfortunately, we don't get a

simple counterpart to proposition 2.9. C_Ω (and therefore Z_Ω, as thus defined) don't play a role relative to $Sp(2n, \mathbb{R})$.

What we end up with is the less-interesting result that $Sp(2n, \mathbb{R})$ and G_Ω stripe $L(V)$ in two distinct ways. When $\Omega = \pm J$ (or, more precisely, when $G_\Omega = Sp(2n, \mathbb{R})$ rather than merely being isomorphic to it), the stripes coincide.

Because the terminology surrounding the symplectic machinery is confusing, we'll conclude with a summary of it:

- "Symplectic form": $(0, 2)$-tensor that is antisymmetric and nondegenerate.
- "Symplectic space (V, ω)": A (necessarily even-dimensional) real vector space, along with a particular symplectic form.
- "Symplectic matrix" (relative to Ω): Given a choice of special antisymmetric, nondegenerate $2n \times 2n$ real matrix Ω, any $2n \times 2n$ real matrix M s.t. $M^T \Omega M = \Omega$.
  This is analogous to "orthogonal matrix", where Id is used as the reference matrix for inner products.
- "Symplectic group" $Sp(2n, \mathbb{R})$: The symplectic matrices when $\Omega = \pm J$, where $J = \begin{pmatrix} 0 & -I_n \\ I_n & 0 \end{pmatrix}$. I.e., $G_{\pm J}$. Every G_Ω (with Ω antisymmetric and nondegenerate) is isomorphic to $Sp(2n, \mathbb{R})$.
  This is analogous to $U(n)$ (or $O(n)$).
- "Symplectic basis" or "canonical basis" (relative to Ω and ω): Given a suitable reference matrix Ω and a particular symplectic form ω, the set of bases in which ω is Ω.
  This is analogous to "orthonormal basis".
- "Symplectic map": A linear map $f : V \to W$ between two symplectic space (V, ω) and (W, ν) s.t. $\nu(f(v), f(v')) = \omega(v, v')$ for all $v, v' \in V$.
  This is analogous to a linear isometry.
- Matrix which can be the representative of a symplectic form: Any antisymmetric, nondegenerate matrix.
- Matrix which can serve as reference matrix Ω: Any antisymmetric, nondegenerate matrix.

2.12. **Aside: Basis Permutations.**

There is a subtlety that we skirted in our discussion of the use of Id as our reference matrix for inner products but which was hinted at by its four standard symplectic counterparts: $\pm J$ and $\pm J'$. We're working with ordered bases, but our basis ordering is arbitrary. Let's briefly examine what happens when we permute the basis.

Let e be a basis for our vector space, let $P(e)$ denote a permuted basis under permutation P, and let S_P be the relevant basis-change matrix from e to $P(e)$.

The form of S_P is easy to infer. It looks like Id but with the rows (or, equivalently, columns) permuted. S_P is all zeroes except a single 1 in each row and column. There are $n!$ such matrices, just as there are $n!$ permutations.

It is obvious that any such matrix S_P is orthogonal, so all the permutations of e are in the same orthogonality class as e. However, if we consider some other $\Omega_0 \neq Id$ — whether as the reference matrix for an inner product and thus symmetric and positive-definite or as the reference matrix for a symplectic form and thus antisymmetric and nondegenerate — S_P need not preserve Ω_0.

Ex. let $\Omega_0 = \begin{pmatrix} 1 & 2 \\ 2 & 8 \end{pmatrix}$ and let $S_P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. $S_P^T \Omega_0 S_P = \begin{pmatrix} 8 & 2 \\ 2 & 1 \end{pmatrix} \neq \Omega_0$.

Since any S_P is real, our discussion applies to any reference matrix for a $(0,2)$-tensor or a sesquilinear form. In all such cases, the object (call it Ω) in question transforms as $\Omega' = (S_P^T)^{-1} \Omega S_P^{-1}$ under such a basis permutation. For S_P to be in the relevant Ω_0-preservation group (what we called G_{Ω_0}), we require $\Omega' = \Omega$, which means $S_P^T \Omega S_P = \Omega$. [Note that for a sesquilinear form, this conditions only applies because S_P is real. In general, the preservation condition is $S^\dagger \Omega S = \Omega$ in that case, and there are complex matrices which satisfy one condition but not the other.]

Consider two bases e and e' that are in the same Ω_0-class. I.e., the relevant basis-change matrix S from e to e' satisfies $S^T \Omega_0 S = \Omega_0$. Now, consider a permutation P. The basis $P(e)$ is obtained from e via S_P and the basis $P(e')$ is obtained from e' via the same S_P, so $P(e')$ is obtained from $P(e)$ via $S_P S S_P^{-1}$.

For $P(e)$ and $P(e')$ to be in the same Ω_0-class, we require that $(S_P S S_P^{-1})^T \Omega_0 (S_P S S_P^{-1}) = \Omega_0$, and there is no reason this should hold. $P(e)$ and $P(e')$ need not be in the same Ω_0-class in general.

However, we see something else. Let $\Omega_0' \equiv S_P \Omega_0 S_P^T$. Then $(S_P S S_P^{-1})^T \Omega_0' (S_P S S_P^{-1}) = S_P S^T S_P^T \Omega_0' S_P S S_P^T$ (since $S_P^{-1} = S_P^T$), which is $S_P S^T S_P^T S_P \Omega_0 S_P^T S_P S S_P^T = S_P S^T \Omega_0 S S_P^T$. Since $S^T \Omega_0 S = \Omega_0$, this is $S_P \Omega_0 S_P^T = \Omega_0'$. I.e., the basis change matrix $S_P S S_P^{-1}$ from $P(e)$ to $P(e')$ preserves Ω_0'.

$S_P \Omega_0 S_P^T$ is a similarity transform (since $S_P^T = S_P^{-1}$) by a real orthogonal matrix. Any similarity transform of Ω_0 preserves its set of eigenvalues, and thus any properties such as nondegeneracy and positive-definiteness as well. Any real orthogonal similarity transform preserves symmetry, antisymmetry, and Hermitianity. Therefore, any valid Ω_0 for the concept of "real inner product", "symplectic form", or "Hermitian inner product" becomes a valid Ω_0' for the same concept.

We now can see what is happening. Under a permutation P, bases in the same Ω_0-class permute to bases in the same $(S_P \Omega_0 S_P^T)$-class (i.e. Ω_0' class). The net effect is to change our choice of reference matrix Ω_0 to Ω_0'. I.e., the world looks the same if we change our choice of reference matrix accordingly. For $\Omega_0 = Id$ this was hidden, because $\Omega_0' = Id$ for all S_P.

Not all valid Ω_0's for one of the aforementioned concepts can be reached from all other Ω_0's this way. The class of all valid reference matrices (which is the same as the class of all possible representative matrices) for a given concept is partitioned into classes within which $\Omega_0' = S_P^T \Omega_0 S_P$ for some S_P. Put another way, the matrices of the form S_P form a subgroup G of $O(n)$ (which is itself a subgroup of $U(n)$ if $K = \mathbb{C}$), and we partition the matrices in the usual manner by $M \sim M'$ iff $M' = S^T M S$ for some $S \in G$.

3. Decomposition of a complex inner product

Any Hermitian inner product can be written as the sum of a real inner product and a symplectic form. Let's see what this means and what it involves. We'll address the details of realification and complexification in a forthcoming set of notes. For now, we'll just provide the bare minimum needed for our present purposes.

Because we'll need to refer to linearity over both $\mathbb{R}$ and $\mathbb{C}$, we'll speak of "real-linear" and "complex-linear" to distinguish between the cases with real and complex coefficients.

Given a set V_S with some $2n$-dimensional real vector space structure on it (which we'll refer to as V_R), along with a real-linear automorphism $\Omega : V_R \to V_R$ s.t. $\Omega^2 = -Id$, we can construct a complex n-dimensional vector space on V_S (which we'll refer to as V_C) that is compatible with V_R in terms of addition and real scalar multiplication and such that Ω implements imaginary multiplication. I.e., $+ : V_R \times V_R \to V_R$ and $\mathbb{R} \times V_R \to V_R$, viewed as operations on V_S, carry over to V_C, and the point $\Omega(v)$ for $v \in V_R$ corresponds to the point iv for that same point v viewed in V_C. It is easy to show that complex scalar multiplication by $z = x + iy$ in V_C then corresponds to the linear operator $xId + y\Omega$ acting on V_R. I.e., we create a faithful field homomorphism from $\mathbb{C}$ to the associative algebra $gl(V_R)$. Different choices of Ω result in different complex vector spaces V_C on the same underlying set as well as different field homomorphisms.

The linear automorphism Ω is often referred to as a "complex structure" on V_R. However, care should be taken with this terminology. "Complex structure" means several things in several other contexts, some compatible or related, and some subtly different.

Going the other way, if we start with a given complex vector space V_C (with underlying set V_S), we can construct a unique compatible real vector space V_R with the same underlying set V_S. Addition and real scalar multiplication restrict to V_R, but the restriction to real scalar multiplication means that we require a larger basis, making V_R $2n$-dimensional. V_C also confers on V_R a natural linear automorphism Ω s.t. $\Omega^2 = -Id$, given by $\Omega(v) = iv$ for every $v \in V_R$ and where v is viewed as an element of V_C on the right. I.e., we convert the trivial complex-linear automorphism $v \to iv$ on V_C to a nontrivial real-linear automorphism $v \to \Omega(v)$ on V_R.

If we wish for other familiar complex functionality — such as notions of the real part, imaginary part, or complex conjugate of a vector — we need to pick a basis e for V_C (or, equivalently, a *compatible* basis for V_R, which most of its bases are not). This is also true if we wish to obtain the standard Hermitian inner product (and norm), which is the pull-back along the corresponding basis map B_e of $\sum z_i \overline{z'_i}$ on the canonical $\mathbb{C}^n$.

I.e., the concepts just mentioned are all basis-dependent. This isn't obvious from casual treatments of complex vector algebra, because they implicitly work in the canonical basis of $\mathbb{C}^n$.

Given V_C and a Hermitian inner product H on it, there is a natural decomposition $H = g + i\omega$ into real and imaginary parts. This decomposition is basis-independent.

Although we cannot divide a vector $v \in V_C$ into real and imaginary parts in a basis-independent way, the situation is different with a tensor or sesquilinear form. $H : V_C \times V_C \to \mathbb{C}$ is a sesquilinear map that produces complex values. We have a canonical way to split a complex scalar into real and imaginary parts (by using the canonical basis for $\mathbb{C}$). I.e., $H(v, w) = g(v, w) + i\omega(v, w)$ for every $v, w \in \mathbb{C}$. Since we are relying on the decomposition of values rather than vectors, the result is patently basis-independent.

g and ω are maps $V_S \times V_S \to \mathbb{R}$ and thus can be viewed as maps $V_R \times V_R \to \mathbb{R}$ or $V_C \times V_C \to \mathbb{R}$ (and they respect addition and real scalar multiplication on both). In the latter capacity, they are not tensors or sesquilinear forms on V_C. However, they *are* $(0, 2)$-tensors on V_R. This is the case for any sesquilinear form H on V_C. Since our H is Hermitian, g is symmetric and ω is antisymmetric. Since H is positive-definite, g is positive-definite and ω is nondegenerate.

We therefore see that, for H a Hermitian inner product on V_C, $H = g + i\omega$, with g a real inner product on V_R and ω a symplectic form on V_R.

It is important to bear in mind that, although each of H, g, and ω can be viewed as a map from $V_R \times V_R$ or $V_C \times V_C$ (since V_R and V_C share the same underlying set), H isn't a tensor on V_R (since it produces complex values) and g and ω aren't tensors or sesquilinear forms on V_C.

The g and ω thus obtained from H are not unrelated or unconstrained. Both are Ω-invariant in the sense that $g(\Omega(v), \Omega(w)) = g(v, w)$ and $\omega(\Omega(v), \Omega(w)) = \omega(v, w)$, and each can be reconstituted from the other, via $g(v, w) = -\omega(v, \Omega(w))$ or any of a number of equivalent relations.

The converse holds too. Any given real inner product g and symplectic form ω on V_R are the real and imaginary parts of some Hermitian inner product on V_C iff $g(v, w) = -\omega(v, \Omega(w))$ and $g(\Omega(v), \Omega(w)) = g(v, w)$. There are other equivalent pairs of identities that suffice too.

g has some important properties. $g(v, v) = H(v, v)$, so the complex norms and real norms of vectors are the same. $g(v, \Omega(v)) = 0$ as well. In fact, the orthonormal bases for V_C relative to H correspond to orthonormal real bases relative to g.

I.e., if $(e_1, \ldots, e_n)$ satisfies $H(e_i, e_j) = \delta_{ij}$, then in the corresponding real basis $(e_1, \ldots, e_n, \Omega(e_1), \ldots, \Omega(e_n))$ for V_R, we have $g(e_i, e_j) = g(\Omega(e_i), \Omega(e_j)) = \delta_{ij}$ and $g(e_i, \Omega(e_j)) = g(\Omega(e_j), e_i) = 0$.

We've omitted all the details and justifications for any of this, because they would lead us too far astray. These will be provided in a forthcoming set of notes on realification and complexification.

4. Complex Angles and Lines

4.1. **Lines through the origin.**

Complex angles and lines can be a bit unintuitive. Consider vector space V over field K. A line through the origin (aka "line", for our purposes) is defined as cv for some $v \neq 0$ and all $c \in K$. We can define a corresponding partition of $V - \{0\}$ via the equivalence relation $v \sim w$ iff $v = cw$ for some $c \neq 0$. The quotient space is just the set of K-lines through the origin. It is known as PV, the projectivization of V over K (with the K-dependence implicit).

Modulo isomorphism, there exists a single vector space over a given K of any given finite dimension n. We therefore can (mostly) rely on our intuition for K^n, where such intuition exists. For the simple case of $\mathbb{R}^n$ we can *mostly* (but not always) rely on our intuition for $\mathbb{R}^2$ and $\mathbb{R}^3$. We know what lines and planes look like in those spaces, so we have a sense of what lines and hyperplanes (but not necessarily general subspaces) look like in $\mathbb{R}^n$. However, this intuition fails for $K = \mathbb{C}$. The smallest vector space ($n = 2$) with a nontrivial notion of a complex line through the origin corresponds to four real dimensions — which is already beyond our capacity for direct visualization.

As mentioned earlier, a complex n-dimensional vector space V_C can be viewed as a real $2n$-dimensional space V_R on the same underlying set, along with a linear automorphism Ω on V_R that implements multiplication by i. A complex line through the origin in V_C is $\{cv; c \in \mathbb{C}\}$ for some $v \neq 0$. This v can also be viewed as a point in V_R (since they share the same underlying set). Since $(xId + y\Omega)v$ in V_R is the same point as $(x + iy)v$ in V_C, the corresponding set of points is $\{(xId + y\Omega)v; x, y \in \mathbb{R}\}$. However $Id(v) = v$, and $\Omega(v)$

is easily seen to be real-linearly independent of v (it corresponds to iv in V_C), so v and $\Omega(v)$ span a real plane in V_R.

A complex line through the origin in V_C is nothing other than a real plane through the origin in V_R. However, the converse does not hold. Not every real plane P through the origin in V_R is a complex line through the origin in V_C. We require that, for some $v \in P$, $\Omega(v) \in P$ too — which turns out to be equivalent to requiring that $\Omega(v) \in P$ for *all* $v \in P$.

I.e. $\Omega(v) \in P$ for some $v \in P$ iff $\Omega(v) \in P$ for *all* $v \in P$. One direction is obvious. To see the other, let $v, w \in P$ and let $\Omega(v) \in P$. Since $\Omega(v)$ is linearly independent of v, the two vectors v and $\Omega(v)$ span P. Therefore, $w = av + b\Omega(v)$ for some real a, b. $\Omega(w) = a\Omega(v) - bv$, since $\Omega^2 = -Id$. Therefore, $\Omega(w)$ is a linear combination of the spanning vectors too and is in P.

(Example of plane in $\mathbb{R}^{2n}$ that isn't a complex line in $\mathbb{C}^n$): Consider the canonical $\mathbb{C}^2$ and $\mathbb{R}^4$ (with their canonical bases), viewed as having the same underlying set (i.e. using the standard map $(x + iy, x' + iy') \to (x, x', y, y')$). In this case, $\Omega = \begin{pmatrix} 0 & -I_2 \\ I_2 & 0 \end{pmatrix}$ (with I_2 the 2×2 identity matrix). Pick $v = (1, 0)$ in $\mathbb{C}^2$, which is $(1, 0, 0, 0)$ in $\mathbb{R}^4$. Then $\Omega(v) = (0, 0, 1, 0)$, which corresponds to $v = (i, 0)$ in $\mathbb{C}^2$. The plane spanned by $(1, 0, 0, 0)$ and $(0, 1, 0, 0)$ does not contain $\Omega(v)$. It corresponds to $\{a(1, 0) + b(0, 1); a, b \in \mathbb{R}\}$ in $\mathbb{C}^2$. This can be written $\{(a, b); a, b \in \mathbb{R}\}$, which is not a complex line. To be a complex line, we need $\{(a + ib)(x + iy, x' + iy'); a, b \in \mathbb{R}\}$ for some $x, x', y, y' \in \mathbb{R}$. This is $\{((ax - by) + i(ay + bx), (ax' - by') + i(ay' + bx')); a, b \in \mathbb{R}\}$. To match our form, we need $ax - by = a$ (i.e. $x = 1$ and $y = 0$) and $ay + bx = 0$ (i.e. $x = y = 0$) and $ax' - by' = b$ (i..e $x' = 0$ and $y' = -1$) and $ay' + bx' = 0$ (i.e. $x' = y' = 0$), which contradict one another. There is no vector $v = (x + iy, x' + iy')$ in $\mathbb{C}^2$ for which our locus is $\{cv; c \in \mathbb{C}\}$.

One word of warning: the term "complex plane" can be used to mean either (i) a plane in V_C (i.e. a subspace of V_C complex-spanned by two vectors), which is how we use it here, or (ii) $\mathbb{C}$ itself (viewed as a real plane), which is how it often appears in complex analysis.

4.2. **Angles.**

Unlike a line (but like a norm), the notion of an angle is not intrinsic to a vector space and requires additional structure. This structure takes the form of an inner product, which defines both norms and angles.

In the real case, we obtain the angle between the lines spanned by nonzero vectors v and w via $g(v, w) = |v||w| \cos\theta$, where $|v| = \sqrt{g(v, v)}$ is the induced norm. We'll take the range of cos^{-1} to be $[0, \pi]$.

Note that we can't choose $[-\pi/2, \pi/2]$ for the range of $\cos^{-1}$ instead, as may be tempting, because cos isn't injective on that domain.

The (possibly obtuse) angle between v and w is $\theta_r = cos^{-1}(g(v, w)/\sqrt{g(v, v)g(w, w)})$. This produces 0 when $v = rw$ for $r > 0$ (i.e. v is parallel to w), $\pi/2$ when $g(v, w) = 0$ (i.e. v and w are orthogonal), and π when $v = rw$ for $r < 0$ (i.e. v is antiparallel to w).

The acute angle between v and w is $\theta_a = cos^{-1}(|g(v, w)/\sqrt{g(v, v)g(w, w)}|)$. This produces 0 when $v = rw$ for any $r \neq 0$ (i.e. v and w span the same line) and is $\pi/2$ when $g(v, w) = 0$.

Despite its $[0, \pi]$ range, $\cos^{-1}$ in the present context can only produce values in the range $[0, \pi/2]$ because of the $||$ inside its argument.

In the complex case, the counterpart of a real inner product $g(-, -)$ is a Hermitian inner product $H(-, -)$. It is less obvious how to extract a notion of angle from such a structure or what it would mean. Let's consider several possible candidates. Without loss of generality, we can assume we are working in an orthonormal basis relative to $H(-, -)$. We saw earlier that $H(v, w)$ on V_C corresponds to $g(v, w) + i\omega(v, w)$,

where g is a real inner product on V_R and ω is a symplectic form on V_R.

As usual, v and w are viewed as points in V_C for $H(v,w)$ and as points in V_R for g and ω, which makes perfect sense since V_R and V_C share the same underlying set. If this expression seems confusing, review section 3.

We can approach the definition of a complex "angle" from a number of ... well ... angles. The simplest would be to demand a real quantity in the range $[0, \pi]$ that is $\pi/2$ for othogonal vectors, 0 for parallel vectors and π for antiparallel vectors.

We could also focus on obtaining an "acute angle", and demand a real quantity in the range $[0, \pi/2]$ that is $\pi/2$ for othogonal vectors and 0 for parallel or antiparallel vectors.

Another approach would be to ask what we we mean by an angle and what we intend to use it for. In a real vector space with an inner product, the norms of vectors and the angles between vectors are invariant under orthonormal basis changes Equivalently, they are invariant under linear isometries of the inner product (i.e. linear automorphisms $f : V \to V$ s.t. $g(f(v), f(w)) = g(v, w)$ for all v, w).

Such isometries preserve norms and angles, but we don't care about norms here. The angle is invariant under a larger class of basis changes — the scaled orthogonal basis changes. Similarly, angles are invariant under a larger class of linear automorphisms $f : V \to V$.

We just need that $g(f(v), f(w))/\sqrt{g(f(v), f(v))g(f(w), f(w))} = g(v, w)/\sqrt{g(v, v)g(w, w)}$.

Such automorphisms form the projective linear group. We won't delve into the properties or details of that group here, since we're just trying to motivate a choice of definition.

Taking this view, we will remain open to the possibility of a complex-valued angle since $H(v, w)$ is complex. Intuitively, we a want a complex angle to have the following features:

- (i) The angle (whether real or complex in value) must be invariant under complex scaling of v and/or w. I.e., we require a meaningful notion of the angle between complex lines, not just complex vectors.
- (ii) The angle must be symmetric. If we swap v and w, we get the same angle.
- (iii) The angle between v and itself (or anything complex-colinear with it) must be 0.
- (iv) We'll ask for the triangle inequality, in the form $|\theta_{v,w}| + |\theta_{w,z}| \geq |\theta_{v,z}|$. However, we'll remain open to relaxing this requirement should it overconstrain us.

  We don't have a natural linear ordering on $\mathbb{C}$, but we don't need one. Because of the $||$, we can define this requirement equally well for complex-valued angles.

We'll see that (i)-(iii) exclude all but one of the obvious candidates, and this remaining candidate satisfies (iv).

Unsurprisingly, the product $|v| \cdot |w|$ will appear in the denominators of all the candidates we consider. An angle can only be meaningful for $v, w \neq 0$, so $|v| = \sqrt{H(v, v)} = \sqrt{g(v, v)}$ and $|w| = \sqrt{H(w, w)} = \sqrt{g(w, w)}$ are both real and positive. The following are some obvious candidates for the angle θ (as a function $\theta(v, w)$ on domain $(V_C - \{0\}) \times (V_C - \{0\})$):

We'll view v and w as vectors in V_C when they appear in $H(v,w)$ and as vectors in V_R when they appear in $g(v,w)$ and $\omega(v,w)$.

Because we only need one example of failure to discount a particular candidate, we lose no generality by considering the canonical $\mathbb{C}^n$ and $\mathbb{R}^{2n}$ and realification $(x_1+iy_1,\dots x_n+iy_n)\to(x_1,\dots,x_n,y_1,\dots,y_n)$ for the purpose of identifying that failure.

- (a) $\cos\theta(v,w) = H(v,w)/(|v|\cdot|w|)$.
- (b) $\cos\theta(v,w) = |H(v,w)|/(|v|\cdot|w|)$.
- (c) $\cos\theta(v,w) = g(v,w)/(|v|\cdot|w|) = (\text{Re } H(v,w))/(|v|\cdot|w|))$.
- (d) $\sin\theta(v,w) = \omega(v,w)/(|v|\cdot|w|) = (\text{Im } H(v,w))/(|v|\cdot|w|))$.
- (e) $\cos\theta(v,w) = |g(v,w)|/(|v|\cdot|w|)$.
- (f) $\sin\theta(v,w) = |\omega(v,w)|/(|v|\cdot|w|)$.
- (g) $\theta(v,w)$ is the acute angle between the real planes in V_R which correspond to the complex lines spanned by v and w in V_C.

Note the $\sin\theta$ (rather than $\cos\theta$) in (d). We need this to make the angle symmetric, since ω is antisymmetric rather than symmetric.

Why not consider definitions with $\sin\theta$ instead of $\cos\theta$ for (b) and (e) and with $\cos\theta$ instead of $\sin\theta$ for (f), since the $||$ puts us in the acute range, where $\cos^{-1}$ and $\sin^{-1}$ are both in $[0,\pi/2]$? We want $\theta(v,v)=0$. This requires $\cos^{-1}(1)$ or $\sin^{-1}(0)$. Since $\omega(v,v)=0$, we can only use $\sin^{-1}$ for it, regardless of the presence of $||$. Similarly, since $H(v,v)=g(v,v)=|v|^2$, we can only use $\cos^{-1}$ for these, regardless of the presence of $||$.

Choice (a) is the only complex-valued candidate angle. Unfortunately, it fails to be invariant under complex scaling of v or w, violating requirement (i). Candidates (c)-(f) also fail to be invariant under complex scaling and violate requirement (i).

Consider (a). If we scale v by $c=re^{i\phi}$, $|v|$ scales by r but $H(v,w)$ (and thus $\cos\theta$) scales by $re^{i\phi}$. Formally, $\cos\theta(cv,w) = e^{i\phi}\cos\theta(v,w)$. This also infects (c), since Re $H(cv,w) = (\cos\phi)(\text{Re } H(v,w)) - (\sin\phi)(\text{Im } H(v,w)) = (\cos\phi)g(v,w) - (\sin\phi)\omega(v,w)$, and (d), since Im $H(cv,w) = (\cos\phi)\omega(v,w) + (\sin\phi)g(v,w)$. (e) and (f) are also implicated, since $|(\cos\phi)g(v,w) - (\sin\phi)\omega(v,w)| \neq |g(v,w)|$ in general, and $|(\cos\phi)\omega(v,w) + (\sin\phi)g(v,w)| \neq |\omega(v,w|)$ in general.

Ex. Pick $w=\Omega(v)$, and consider (a). $H(v,w) = g(v,\Omega(v)) + i\omega(v,\Omega(v)) = 0 - i|v|^2$. Since $g(v,v) = g(\Omega(v),\Omega(v))$, $|v| = |w|$. Therefore, $\cos\theta(v,w) = -i$ and $\cos\theta(cv,w) = -ie^{i\phi} = \sin\phi - i\cos\phi$. Under definition (d), we can read off the imaginary parts above to get $\sin\theta(v,w) = -1$ but $\sin\theta(cv,w) = -\cos\phi$. Under definition (c), we can read off the real parts above to get $\cos\theta(v,w)=0$ but $\cos\theta(cv,w)=\sin\phi$. Taking the magnitudes doesn't help. We can read off (f) as $\sin\theta(v,w)=1$ and $\sin\theta(cv,w)=|\cos\phi|$. Similarly, we can read off (e) as $\cos\theta(v,w)=0$ and $\cos\theta(cv,w)=|\sin\phi|$. In all these cases, we fail to be invariant under complex-scaling.

Let's next consider (g). We want to compute the angle between two 2-dimensional subspaces (aka 2-planes or real planes) in V_R. If v and w are the vectors in V_C that span the relevant complex lines, then v and $\Omega(v)$ span one real plane, and w and $\Omega(w)$ span the other. Moreover, of the 6 pairs of vectors obtained from $(v,w,\Omega(v),\Omega(w))$, all but two are automatically orthogonal under g. The two which need not (but still can) be orthogonal are (v,w) and $(\Omega(v),\Omega(w))$.

Since $g(\Omega(v),\Omega(w)) = g(v,w)$, one pair is orthogonal iff the other is.

There are two standard ways to calculate the angle we want. (i) We can calculate the Jordan principal angles or (ii) we can use the induced inner product on the exterior algebra over V_R.

The Jordan approach is simple to state but can be difficult to calculate with. Between a given k-dimensional subspace W and a given m-dimensional subspace W' of some real n-dimensional vector space V equipped with a real inner product g, there are $\min(k,m)$ Jordan principal angles, and these form a nondecreasing sequence with values in $[0,\pi/2]$.

The first angle θ_1 is calculated by taking the minimum angle (via the usual $\cos\theta = g(w,w')/\sqrt{g(w,w)g(w',w')}$ formula) between any $w \in W$ and $w' \in W'$. We then pick a pair of vectors (one from each subspace) that realizes this minimum. The second angle θ_2 is also the minimum acute angle, but now restricted to vectors in W and W' orthogonal to the chosen one in each subspace. We then repeat the process until we run out of dimensions (hence the $\min(k,m)$).

In our case, $k = m = 2$, and n becomes $2n$, so there are two Jordan principal angles. We are also subject to the orthogonality constraints mentioned above, and these simplify the calculation considerably. We won't work through the details here, but it is not terribly difficult. The two principal angles turn out to be the same, and they are equal to definition (b).

The details of the calculation will be provided in a forthcoming set of notes on the angle between subspaces of $\mathbb{R}^n$.

The exterior algebra approach is a simple case of a more general method. Given a k-dimensional subspace W and an m-dimensional subspace W' of a real n-dimensional space V equipped with an inner product g, we can compute an angle between W and W' via an induced inner-product-like object on the Clifford algebra associated with g over V.

However, our case does not require this level of generality. When $k = m$, we need look no further than the exterior algebra. g induces a true inner product $\tilde{g}$ on the exterior algebra of V, which is itself a vector space over $\mathbb{R}$. This inner product is trivial between grades (i.e. k-blades and m-blades are orthogonal when $k \neq m$), but within a given grade it has substance. It therefore is useful only when $k = m$.

Recall that a k-blade is of the form $\wedge_{i=1}^{k} v_i$ for some set of vectors $v_i \in V$.

Since a k-blade spans a k-space, and we can use $\tilde{g}$ to compute an angle between k-blades (via the usual $\cos\theta = \tilde{g}(A,B)/\sqrt{\tilde{g}(A,A)\tilde{g}(B,B)}$, where A and B are k-blades viewed as vectors in $\Lambda^k V$), we can view the latter as an angle between the corresponding k-spaces. It is easy to show that this angle is independent of the particular choice of k-blades that span the k-spaces in question.

In our case, $k = m = 2$, n becomes $2n$, and the blades in question are $v \wedge \Omega(v)$ and $w \wedge \Omega(w)$. The aforementioned orthogonality constraints also come into play. Once again, we won't perform the calculation here, and we refer the curious reader to our forthcoming set of notes. The result turns out to be the same as definition (b).

One aspect of our discussion may seem odd. Definition (b) is framed in terms of $|H|$, and clearly depends on both g and ω. On the other hand, the Jordan and exterior algebra approaches know only about g, yet somehow reproduce (b). How is it possible that they restore the extra information from ω? The answer is something we alluded to earlier. The g and ω induced by H are not independent. In fact, each can be derived from the other (and they are further constrained to be Ω-invariant). I.e., the entire information content of H is present in each of g and ω.

We now have narrowed the obvious candidates for a definition of complex angle to a single choice: (b). It turns out to satisfy all our demands, including the potentially optional (iv).

(i) The $||$ eliminates the extra $e^{i\phi}$ factor that plagued (a) and (c)-(f).

(ii) The denominator is symmetric. $|H(w,v)| = |\overline{H(v,w)}| = |H(v,w)|$, so we have symmetry.

(iii) The $H(v,v) = |v|^2$, so $\cos\theta = 1$ and $\theta = 0$.

(iv) This follows immediately from the exterior algebra approach. Since the $\theta(v,w)$ of (b) is also that obtained from the exterior algebra approach, and the latter is just the usual angle between real vectors (albeit in $\Lambda^2 V_R$), it follows that the angles automatically satisfy the triangle inequality, like in any other real vector space.

In conclusion, of the obvious choices, we have a single viable candidate for an angle between complex lines. It produces real values. This angle is $\cos\theta(v,w) = |H(v,w)|/(|v| \cdot |w|)$, and it is always acute (i.e. $\theta(v,w) \in [0, \pi/2]$), since the right side is nonnegative.

The angle thus defined is sometimes referred to as the "quantum angle", because it appears in quantum mechanics.

This is not surprising. The true state space of a quantum mechanical system is a projective Hilbert space, not the Hilbert space itself. A Hilbert space is a complex vector space equipped with a Hermitian inner product (and which is complete relative to the metric induced by that inner product). The quantum angle is the part of the Hermitian inner product that survives projectivization. It is the *only* part that does so.