Notes on Elementary Linear Algebra

Adam Coffman

Contents

These Notes are compiled from classroom handouts for Math 351, Math 511, and Math 554 at IPFW. They are not self-contained, but supplement the required texts, [A], [FIS], and [HK].

1 Real vector spaces

Definition 1.1. Given a set V, and two operations $+$ (addition) and $\cdot$ (scalar multiplication), V is a "real vector space" means that the operations have all of the following properties:

1. Closure under Addition: For any $\mathbf{u} \in V$ and $\mathbf{v} \in V$, $\mathbf{u} + \mathbf{v} \in V$.
2. Associative Law for Addition: For any $\mathbf{u} \in V$ and $\mathbf{v} \in V$ and $\mathbf{w} \in V$, $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$.
3. Existence of a Zero Element: There exists an element $\mathbf{0} \in V$ such that for any $\mathbf{v} \in V$, $\mathbf{v} + \mathbf{0} = \mathbf{v}$.
4. Existence of an Opposite: For each $\mathbf{v} \in V$, there exists an element of V, called $-\mathbf{v} \in V$, such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.
5. Closure under Scalar Multiplication: For any $r \in \mathbb{R}$ and $\mathbf{v} \in V$, $r \cdot \mathbf{v} \in V$.
6. Associative Law for Scalar Multiplication: For any $r, s \in \mathbb{R}$ and $\mathbf{v} \in V$, $(rs) \cdot \mathbf{v} = r \cdot (s \cdot \mathbf{v})$.
7. Scalar Multiplication Identity: For any $\mathbf{v} \in V$, $1 \cdot \mathbf{v} = \mathbf{v}$.
8. Distributive Law: For all $r, s \in \mathbb{R}$ and $\mathbf{v} \in V$, $(r + s) \cdot \mathbf{v} = (r \cdot \mathbf{v}) + (s \cdot \mathbf{v})$.
9. Distributive Law: For all $r \in \mathbb{R}$ and $\mathbf{u}, \mathbf{v} \in V$, $r \cdot (\mathbf{u} + \mathbf{v}) = (r \cdot \mathbf{u}) + (r \cdot \mathbf{v})$.

The following theorems refer to a real vector space V. Theorems 1.2 through 1.11 use only the first four axioms about addition.

Theorem 1.2 (Right Cancellation)**.** *Given* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$*, if* $\mathbf{u} + \mathbf{w} = \mathbf{v} + \mathbf{w}$*, then* $\mathbf{u} = \mathbf{v}$*.*

Proof. $\mathbf{u} + \mathbf{w}$ and $\mathbf{v} + \mathbf{w}$ are elements of V by Axiom 1. Since $\mathbf{w} \in V$, there exists an opposite, also called an "additive inverse," $-\mathbf{w} \in V$. Adding this to both sides of $\mathbf{u} + \mathbf{w} = \mathbf{v} + \mathbf{w}$ on the right gives $(\mathbf{u} + \mathbf{w}) + (-\mathbf{w}) = (\mathbf{v} + \mathbf{w}) + (-\mathbf{w})$, and the associative law gives $\mathbf{u} + (\mathbf{w} + (-\mathbf{w})) = \mathbf{v} + (\mathbf{w} + (-\mathbf{w}))$, so $\mathbf{u} + \mathbf{0} = \mathbf{v} + \mathbf{0}$. By Axiom 3, $\mathbf{u} = \mathbf{v}$. ∎

Theorem 1.3. *Given* $\mathbf{u}, \mathbf{w} \in V$*, if* $\mathbf{u} + \mathbf{w} = \mathbf{w}$*, then* $\mathbf{u} = \mathbf{0}$*.*

Proof. Since $\mathbf{w} \in V$, there exists an additive inverse $-\mathbf{w} \in V$. Adding this to both sides of $\mathbf{u} + \mathbf{w} = \mathbf{w}$ on the right gives $(\mathbf{u} + \mathbf{w}) + (-\mathbf{w}) = \mathbf{w} + (-\mathbf{w})$, and the associative law gives $\mathbf{u} + (\mathbf{w} + (-\mathbf{w})) = \mathbf{w} + (-\mathbf{w})$, so $\mathbf{u} + \mathbf{0} = \mathbf{0}$. By Axiom 3, $\mathbf{u} = \mathbf{0}$. ∎

Theorem 1.4. *For any* $\mathbf{v} \in V$*,* $(-\mathbf{v}) + \mathbf{v} = \mathbf{0}$*.*

Proof. $(-\mathbf{v}) + \mathbf{v} \in V$ by Axiom 1. The following steps use Axioms 2, 3, 4.

$$\begin{aligned}
((-\mathbf{v}) + \mathbf{v}) + ((-\mathbf{v}) + \mathbf{v}) &= (((-\mathbf{v}) + \mathbf{v}) + (-\mathbf{v})) + \mathbf{v} \\
&= ((-\mathbf{v}) + (\mathbf{v} + (-\mathbf{v}))) + \mathbf{v} \\
&= ((-\mathbf{v}) + \mathbf{0}) + \mathbf{v} \\
&= (-\mathbf{v}) + \mathbf{v},
\end{aligned}$$

so the previous Theorem applies with $\mathbf{u}$ and $\mathbf{w}$ both equal to $(-\mathbf{v}) + \mathbf{v}$, to show $(-\mathbf{v}) + \mathbf{v} = \mathbf{0}$. ∎

Theorem 1.5. *For any* $\mathbf{v} \in V$*,* $\mathbf{0} + \mathbf{v} = \mathbf{v}$*.*

Proof. We use the fact that $\mathbf{v}$ has an additive inverse, the associative law, and the previous Theorem.

$$\mathbf{0} + \mathbf{v} = (\mathbf{v} + (-\mathbf{v})) + \mathbf{v} = \mathbf{v} + ((-\mathbf{v}) + \mathbf{v}) = \mathbf{v} + \mathbf{0} = \mathbf{v}.$$

∎

Theorem 1.6 (Left Cancellation)**.** *Given* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$*, if* $\mathbf{w} + \mathbf{u} = \mathbf{w} + \mathbf{v}$*, then* $\mathbf{u} = \mathbf{v}$*.*

Proof. $\mathbf{w} + \mathbf{u}$ and $\mathbf{w} + \mathbf{v}$ are in V by Axiom 1. Since $\mathbf{w} \in V$, there exists an additive inverse $-\mathbf{w} \in V$. Adding this to both sides of $\mathbf{w} + \mathbf{u} = \mathbf{w} + \mathbf{v}$ on the left gives $(-\mathbf{w}) + (\mathbf{w} + \mathbf{u}) = (-\mathbf{w}) + (\mathbf{w} + \mathbf{v})$, and the associative law gives $((-\mathbf{w}) + \mathbf{w}) + \mathbf{u} = ((-\mathbf{w}) + \mathbf{w}) + \mathbf{v}$. By Theorem 1.4, $\mathbf{0} + \mathbf{u} = \mathbf{0} + \mathbf{v}$, and by the previous Theorem, $\mathbf{u} = \mathbf{v}$. ∎

Theorem 1.7 (Uniqueness of Zero Element)**.** *Given* $\mathbf{u}, \mathbf{w} \in V$*, if* $\mathbf{w} + \mathbf{u} = \mathbf{w}$*, then* $\mathbf{u} = \mathbf{0}$*.*

Proof. $\mathbf{w} = \mathbf{w} + \mathbf{0}$ by Axiom 3, so if $\mathbf{w} + \mathbf{u} = \mathbf{w}$, then $\mathbf{w} + \mathbf{u} = \mathbf{w} + \mathbf{0}$, and the previous Theorem gives $\mathbf{u} = \mathbf{0}$. ∎

Theorem 1.8 (Uniqueness of Additive Inverse)**.** *Given* $\mathbf{v}, \mathbf{w} \in V$*, if* $\mathbf{v} + \mathbf{w} = \mathbf{0}$ *then* $\mathbf{v} = -\mathbf{w}$ *and* $\mathbf{w} = -\mathbf{v}$*.*

Proof. $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$ by Axiom 4, so if $\mathbf{v} + \mathbf{w} = \mathbf{0}$, then $\mathbf{v} + \mathbf{w} = \mathbf{v} + (-\mathbf{v})$, and the Left Cancellation theorem gives $\mathbf{w} = -\mathbf{v}$.

$(-\mathbf{w}) + \mathbf{w} = \mathbf{0}$ by Theorem 1.4, so if $\mathbf{v} + \mathbf{w} = \mathbf{0}$, then $\mathbf{v} + \mathbf{w} = (-\mathbf{w}) + \mathbf{w}$, and the Right Cancellation theorem gives $\mathbf{v} = -\mathbf{w}$. ∎

Theorem 1.9. $-\mathbf{0} = \mathbf{0}$*.*

Proof. $\mathbf{0} + \mathbf{0} = \mathbf{0}$ by Axiom 3, so the previous Theorem applies, with $\mathbf{v} = \mathbf{0}$ and $\mathbf{w} = \mathbf{0}$, to show that $\mathbf{0} = -\mathbf{0}$. ∎

Theorem 1.10. *For any* $\mathbf{v} \in V$*,* $-(-\mathbf{v}) = \mathbf{v}$*.*

Proof. Since $-\mathbf{v} \in V$ and $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$ by Axiom 4, Theorem 1.8 applies, with $\mathbf{w} = -\mathbf{v}$, to show $\mathbf{v} = -\mathbf{w} = -(-\mathbf{v})$. ∎

Theorem 1.11. *Given* $\mathbf{u}, \mathbf{x} \in V$*,* $-(\mathbf{u} + \mathbf{x}) = (-\mathbf{x}) + (-\mathbf{u})$*.*

Proof. Note $-\mathbf{x}$ and $-\mathbf{u}$ are in V by Axiom 4, and $\mathbf{u} + \mathbf{x}$ and $(-\mathbf{x}) + (-\mathbf{u})$ are in V by Axiom 1. Consider the sum $(\mathbf{u} + \mathbf{x}) + ((-\mathbf{x}) + (-\mathbf{u}))$. Using the associative law, it simplifies: $\mathbf{u}+(\mathbf{x}+((-\mathbf{x})+(-\mathbf{u}))) = \mathbf{u}+((\mathbf{x}+(-\mathbf{x}))+(-\mathbf{u})) = \mathbf{u}+(\mathbf{0}+(-\mathbf{u})) = (\mathbf{u}+\mathbf{0})+(-\mathbf{u}) = \mathbf{u}+(-\mathbf{u}) = \mathbf{0}$. So, Theorem 1.8 applies, with $\mathbf{v} = \mathbf{u} + \mathbf{x}$ and $\mathbf{w} = (-\mathbf{x}) + (-\mathbf{u})$, to show $\mathbf{w} = -\mathbf{v}$, and $(-\mathbf{x}) + (-\mathbf{u}) = -(\mathbf{u} + \mathbf{x})$. ∎

The previous results only used Axioms 1 – 4, about "+," but the next result, even though its statement refers only to +, uses a scalar multiplication trick, together with the distributive axioms, which relate scalar multiplication to addition.

Theorem 1.12 (Commutative Property of Addition)**.** *For any* $\mathbf{v}, \mathbf{w} \in V$*,* $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$*.*

Proof. We start with $(1+1) \cdot (\mathbf{v} + \mathbf{w})$, which is in V by both closure axioms, set LHS=RHS, and use both distributive laws:

$$\begin{aligned}
(1+1)\cdot(\mathbf{v}+\mathbf{w}) &= (1+1)\cdot(\mathbf{v}+\mathbf{w}) \\
((1+1)\cdot\mathbf{v}) + ((1+1)\cdot\mathbf{w}) &= (1\cdot(\mathbf{v}+\mathbf{w})) + (1\cdot(\mathbf{v}+\mathbf{w})) \\
((1\cdot\mathbf{v}) + (1\cdot\mathbf{v})) + ((1\cdot\mathbf{w}) + (1\cdot\mathbf{w})) &= (\mathbf{v}+\mathbf{w}) + (\mathbf{v}+\mathbf{w}) \\
(\mathbf{v}+\mathbf{v}) + (\mathbf{w}+\mathbf{w}) &= (\mathbf{v}+\mathbf{w}) + (\mathbf{v}+\mathbf{w}).
\end{aligned}$$

Then, the associative law gives $\mathbf{v} + (\mathbf{v} + (\mathbf{w} + \mathbf{w})) = \mathbf{v} + (\mathbf{w} + (\mathbf{v} + \mathbf{w}))$, and Left Cancellation leaves $\mathbf{v} + (\mathbf{w} + \mathbf{w}) = \mathbf{w} + (\mathbf{v} + \mathbf{w})$. Using the associative law again, $(\mathbf{v} + \mathbf{w}) + \mathbf{w} = (\mathbf{w} + \mathbf{v}) + \mathbf{w}$, and Right Cancellation gives the result $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$. ∎

Theorem 1.13. *For any* $\mathbf{v} \in V$*,* $0 \cdot \mathbf{v} = \mathbf{0}$*.*

Proof. $0 \cdot \mathbf{v} \in V$ by Axiom 5. The distributive law is needed. $0 \cdot \mathbf{v} = (0+0) \cdot \mathbf{v} = (0 \cdot \mathbf{v}) + (0 \cdot \mathbf{v})$. Theorem 1.3 applies, with $\mathbf{u}$ and $\mathbf{w}$ both equal to $0 \cdot \mathbf{v}$, to show $0 \cdot \mathbf{v} = \mathbf{0}$. ∎

Theorem 1.14. *For any* $\mathbf{v} \in V$*,* $(-1) \cdot \mathbf{v} = -\mathbf{v}$*.*

Proof. $(-1) \cdot \mathbf{v} \in V$ by Axiom 5. Using Axiom 7, the distributive law, and the previous Theorem, $\mathbf{v} + ((-1) \cdot \mathbf{v}) = (1 \cdot \mathbf{v}) + ((-1) \cdot \mathbf{v}) = (1 + (-1)) \cdot \mathbf{v} = 0 \cdot \mathbf{v} = \mathbf{0}$. Theorem 1.8 applies, with $\mathbf{w} = (-1) \cdot \mathbf{v}$, to show $-\mathbf{v} = \mathbf{w} = (-1) \cdot \mathbf{v}$. ∎

Theorem 1.15. *For any* $r \in \mathbb{R}$*,* $r \cdot \mathbf{0} = \mathbf{0}$*.*

Proof. $r \cdot \mathbf{0} \in V$ by Axiom 5. Using the distributive law, $r \cdot \mathbf{0} = r \cdot (\mathbf{0} + \mathbf{0}) = (r \cdot \mathbf{0}) + (r \cdot \mathbf{0})$. Theorem 1.3 applies with $\mathbf{u} = \mathbf{w} = r \cdot \mathbf{0}$, to show $r \cdot \mathbf{0} = \mathbf{0}$. ∎

Theorem 1.16. *For any* $r \in \mathbb{R}$ *and* $\mathbf{u} \in V$*,* $(-r) \cdot \mathbf{u} = -(r \cdot \mathbf{u})$*.*

Proof. $(-r) \cdot \mathbf{u}$ and $r \cdot \mathbf{u}$ are in V by Axiom 5. Using the distributive law, and Theorem 1.13, $(r \cdot \mathbf{u}) + ((-r) \cdot \mathbf{u}) = (r + (-r)) \cdot \mathbf{u} = 0 \cdot \mathbf{u} = \mathbf{0}$. Theorem 1.8 applies, with $\mathbf{v} = r \cdot \mathbf{u}$ and $\mathbf{w} = (-r) \cdot \mathbf{u}$, to show $\mathbf{w} = -\mathbf{v}$, so $(-r) \cdot \mathbf{v} = -(r \cdot \mathbf{v})$. ∎

Theorem 1.17. *Given* $r \in \mathbb{R}$ *and* $\mathbf{u} \in V$*, if* $r \cdot \mathbf{u} = \mathbf{0}$*, then* $r = 0$ *or* $\mathbf{u} = \mathbf{0}$*.*

Proof. There are two cases: given $r \in \mathbb{R}$, either $r = 0$, in which case the Theorem is proved already, or $r \neq 0$. So, supposing $r \neq 0$, multiply both sides of $r \cdot \mathbf{u} = \mathbf{0}$ by $\frac{1}{r}$, to get $\frac{1}{r} \cdot (r \cdot \mathbf{u}) = \frac{1}{r} \cdot \mathbf{0}$. By Axioms 6 and 7, the LHS simplifies to $(\frac{1}{r} r) \cdot \mathbf{u} = 1 \cdot \mathbf{u} = \mathbf{u}$, and by Theorem 1.15, the RHS simplifies to $\mathbf{0}$, proving $\mathbf{u} = \mathbf{0}$. ∎

Theorem 1.18. *For any* $\mathbf{v} \in V$*, the following are equivalent: (1)* $\mathbf{v} + \mathbf{v} = \mathbf{0}$*, (2)* $\mathbf{v} = -\mathbf{v}$*, (3)* $\mathbf{v} = \mathbf{0}$.

Proof. (1) $\implies$ (2) by Theorem 1.8. To show (2) $\implies$ (1), start with $\mathbf{v} = -\mathbf{v}$ and add $\mathbf{v}$ to both sides on the left to get $\mathbf{v} + \mathbf{v} = \mathbf{v} + (-\mathbf{v}) = \mathbf{0}$. (3) $\implies$ (1) just by Axiom 3: $\mathbf{0} + \mathbf{0} = \mathbf{0}$, so if $\mathbf{v} = \mathbf{0}$, then $\mathbf{v} + \mathbf{v} = \mathbf{0}$. Note that so far, we have only used the axioms and theorems for addition, but to show (1) $\implies$ (3), which establishes the equivalences of the Theorem, we need properties of scalar multiplication. If $\mathbf{0} = \mathbf{v} + \mathbf{v}$, then $\mathbf{0} = (1 \cdot \mathbf{v}) + (1 \cdot \mathbf{v}) = (1 + 1) \cdot \mathbf{v}$. Theorem 1.17 applies, and since $(1 + 1) \neq 0 \in \mathbb{R}$, $\mathbf{v}$ must be $\mathbf{0}$. ∎

Definition 1.19. It is convenient to abbreviate the sum $\mathbf{v} + (-\mathbf{w})$ as $\mathbf{v} - \mathbf{w}$. This defines vector subtraction, so that "$\mathbf{v}$ minus $\mathbf{w}$" is defined to be the sum of $\mathbf{v}$ and the additive inverse of $\mathbf{w}$.

Notation 1.20. Considering the associative law for addition, it is convenient to write the sum of more than two terms without all the parentheses: $\mathbf{u} + \mathbf{v} + \mathbf{w}$ can mean either $(\mathbf{u} + \mathbf{v}) + \mathbf{w}$, or $\mathbf{u} + (\mathbf{v} + \mathbf{w})$, since we get the same result either way. In light of Theorem 1.16, we can write $-r \cdot \mathbf{v}$ to mean either $(-r) \cdot \mathbf{v}$ or $-(r \cdot \mathbf{v})$, since these are the same. We can also drop the "dot" for scalar multiplication, when it is clear which symbols are scalars and which are vectors: instead of $3 \cdot \mathbf{u}$, just write $3\mathbf{u}$. It is also convenient to establish an "order of operations," so that just like with arithmetic, scalar multiplication is done before addition or subtraction. So, $4\mathbf{v} + \mathbf{u} - 3\mathbf{w}$ is a short way to write $(4 \cdot \mathbf{v}) + (\mathbf{u} + (-(3 \cdot \mathbf{w})))$.

2 Subspaces

The general idea of the statement "W is a subspace of V" is that W is a vector space contained in a bigger vector space V, and the $+$ and $\cdot$ operations are the same in W as they are in V.

Definition 2.1. Let $(V, +_V, \cdot_V)$ be a vector space. A set W is called a subspace of V means:

- $W \subseteq V$, and
- There are operations $+_W$ and $\cdot_W$ such that $(W, +_W, \cdot_W)$ is a real vector space, and
- For all $\mathbf{x}, \mathbf{y} \in W$, $\mathbf{x} +_V \mathbf{y} = \mathbf{x} +_W \mathbf{y}$, and
- For all $\mathbf{x} \in W$, $r \in \mathbb{R}$, $r \cdot_V \mathbf{x} = r \cdot_W \mathbf{x}$.

Theorem 2.2. *If W is a subspace of V, where V has zero element 0_V, then 0_V is an element of W, and is equal to the zero element of W.*

Proof. By the second part of Definition 2.1, W is a vector space, so by Property 3. of Definition 1.1 applied to W, W contains a zero element $0_W \in W$. By the first part of Definition 2.1, $W \subseteq V$, which implies $0_W \in V$. By part 3. of Definition 1.1 applied to W, $0_W +_W 0_W = 0_W$, and by Definition 2.1, $0_W +_V 0_W = 0_W +_W 0_W$. It follows that $0_W +_V 0_W = 0_W \in V$, and then Theorem 1.3 implies $0_W = 0_V$. ∎

Theorem 2.2 can be used in this way: if W is a set that does not contain 0_V as one of its elements, then W is not a subspace of V.

Theorem 2.3. *If W is a subspace of V, then for every $\mathbf{w} \in W$, the opposite of $\mathbf{w}$ in W is the same as the opposite of $\mathbf{w}$ in V.*

Proof. Let $\mathbf{w}$ be an element of W; then $\mathbf{w} \in V$ because $W \subseteq V$.

First, we show that an additive inverse of $\mathbf{w}$ in W is also an additive inverse of $\mathbf{w}$ in V. Let $\mathbf{y}$ be any additive inverse of $\mathbf{w}$ in W, meaning $\mathbf{y} \in W$ and $\mathbf{w} +_W \mathbf{y} = 0_W$. (There exists at least one such $\mathbf{y}$, by Definition 1.1 applied to W.) $W \subseteq V$ implies $\mathbf{y} \in V$. From Theorem 2.2, $0_W = 0_V$, and $\mathbf{w} +_W \mathbf{y} = \mathbf{w} +_V \mathbf{y}$ by Definition 2.1, so $\mathbf{w} +_V \mathbf{y} = 0_V$, which means $\mathbf{y}$ is an additive inverse of $\mathbf{w}$ in V.

Second, we show that an additive inverse of $\mathbf{w}$ in V is also an additive inverse of $\mathbf{w}$ in W. Let $\mathbf{z}$ be any additive inverse of $\mathbf{w}$ in V, meaning $z \in V$ and $\mathbf{w} +_V \mathbf{z} = 0_V$. (There exists at least one such $\mathbf{z}$, by Definition 1.1 applied to V.) Then $\mathbf{w} +_V \mathbf{z} = 0_V = \mathbf{w} +_V \mathbf{y}$, so by Left Cancellation in V, $\mathbf{z} = \mathbf{y}$ and $\mathbf{y} \in W$, which imply $\mathbf{z} \in W$ and $\mathbf{w} +_W \mathbf{z} = \mathbf{w} +_W \mathbf{y} = 0_W$, meaning $\mathbf{z}$ is an additive inverse of $\mathbf{w}$ in W.

By uniqueness of opposites (Theorem 1.8 applied to either V or W), we can refer to $\mathbf{y} = \mathbf{z}$ as "the" opposite of $\mathbf{w}$, and denote it $\mathbf{y} = -\mathbf{w}$. ∎

Theorem 2.3 also implies that subtraction in W is the same as subtraction in V: by Definition 1.19, for $\mathbf{v}, \mathbf{w} \in W$, $\mathbf{v} -_W \mathbf{w} = \mathbf{v} +_W \mathbf{y} = \mathbf{v} +_V \mathbf{y} = \mathbf{v} -_V \mathbf{w}$.

Theorem 2.3 can be used in this way: if W is a subset of a vector space V and there is an element $\mathbf{w} \in W$, where the opposite of $\mathbf{w}$ in V is not an element of W, then W is not a subspace of V.

Theorem 2.4. *Let $(V, +_V, \cdot_V)$ be a real vector space, and let W be a subset of V. Then W, with the same addition and scalar multiplication operations, is a subspace of V if and only if:*

(1) $\mathbf{x} \in W$, $\mathbf{y} \in W$ imply $\mathbf{x} +_V \mathbf{y} \in W$ (closure under $+_V$ addition), and

(2) $r \in \mathbb{R}$, $\mathbf{x} \in W$ imply $r \cdot_V \mathbf{x} \in W$ (closure under $\cdot_V$ scalar multiplication), and

(3) $W \neq \emptyset$.

Proof. Let V have zero element 0_V.

First suppose W is a subspace, so that as in the Proof of Theorem 2.2, W contains a zero element 0_W, which shows $W \neq \emptyset$, and (3) is true. From Property 1. of Definition 1.1, $\mathbf{x} \in W$, $\mathbf{y} \in W$ imply $\mathbf{x} +_W \mathbf{y} \in W$, and from the definition of subspace, $\mathbf{x} +_W \mathbf{y} = \mathbf{x} +_V \mathbf{y}$, so $\mathbf{x} +_V \mathbf{y} \in W$, establishing (1). Similarly, from Property 5. of Definition 1.1, $r \in \mathbb{R}$ implies $r \cdot_W \mathbf{x} \in W$, and from the definition of subspace, $r \cdot_W \mathbf{x} = r \cdot_V \mathbf{x}$, so $r \cdot_V \mathbf{x} \in W$, establishing (2).

Conversely, it follows from (1), (2), and (3) that W is a subspace of V, as follows: W is a subset of V by hypothesis. Define $+_W$ and $\cdot_W$ by $\mathbf{x} +_W \mathbf{y} = \mathbf{x} +_V \mathbf{y}$, and $r \cdot_W \mathbf{x} = r \cdot_V \mathbf{x}$ — these define operations on W by (1) and (2) (the closure Properties 1. and 5. from Definition 1.1, and also parts of Definition 2.1), but it remains to check the other properties to show that $(W, +_W, \cdot_W)$ is a vector space. Since $W \neq \emptyset$ by (3), there is some $\mathbf{x} \in W$, and by (2), $0 \cdot_V \mathbf{x} \in W$. By Theorem 1.13, $0 \cdot_V \mathbf{x} = 0_V$, so $0_V \in W$, and it satisfies $\mathbf{x} +_W 0_V = \mathbf{x} +_V 0_V = \mathbf{x}$ for all $\mathbf{x} \in W$, so 0_V is a zero element for W. The scalar multiple identity also works: $1 \cdot_W \mathbf{x} = 1 \cdot_V \mathbf{x} = \mathbf{x}$. Also by (2), for any $\mathbf{x} \in W$, $(-1) \cdot_V \mathbf{x} \in W$, and it is easy to check $(-1) \cdot_V \mathbf{x}$ is an additive inverse of $\mathbf{x}$ in W: $\mathbf{x} +_W ((-1) \cdot_V \mathbf{x}) = (1 \cdot_V \mathbf{x}) +_V ((-1) \cdot_V \mathbf{x}) = (1 + (-1)) \cdot_V \mathbf{x} = 0 \cdot_V \mathbf{x} = 0_V$. The other vector space properties, (2,6,8,9) from Definition 1.1, follow immediately from the facts that these properties hold in V and the operations in W give the same sums and scalar multiples. ∎

3 Additive Functions and Linear Functions

Let U and V be real vector spaces.

Definition 3.1. A function $f : U \to V$ is "additive" means: f has the property that $f(\mathbf{a}+\mathbf{b}) = f(\mathbf{a}) + f(\mathbf{b})$ for all $\mathbf{a}, \mathbf{b} \in U$.

Definition 3.2. A function $f : U \to V$ is "linear" means: f is additive and also has the "scaling property": $f(r \cdot \mathbf{a}) = r \cdot f(\mathbf{a})$ for all $r \in \mathbb{R}$ and all $\mathbf{a} \in U$.

Exercise 3.3. If $f : U \to V$ is additive, then: $f(\mathbf{0}_U) = \mathbf{0}_V$, and for all $\mathbf{a} \in U$, $f(-\mathbf{a}) = -f(\mathbf{a})$.

Lemma 3.4. *If $f : U \to V$ is additive, then for every integer $n \in \mathbb{Z}$, $f(n \cdot \mathbf{a}) = n \cdot f(\mathbf{a})$.*

Proof. The $n = 0$ case follows from Exercise 3.3 and vector space properties. The $n = 1$ case follows from the vector space axiom for $1 \cdot \mathbf{a}$. If the claim holds for a positive integer n, then it also holds for the negative integer $-n$: $f((-n)\cdot\mathbf{a}) = f(-(n\cdot\mathbf{a})) = -f(n\cdot\mathbf{a}) = -(n\cdot f(\mathbf{a})) = (-n)\cdot f(\mathbf{a})$, using Exercise 3.3 and Theorem 1.16.

To prove the claim for positive integers by induction on n, suppose $f(n \cdot \mathbf{a}) = n \cdot f(\mathbf{a})$, and we want to show $f((n+1) \cdot \mathbf{a}) = (n+1) \cdot f(\mathbf{a})$.

$$f((n+1) \cdot \mathbf{a}) = f((n \cdot \mathbf{a}) + (1 \cdot \mathbf{a})) = f(n \cdot \mathbf{a}) + f(1 \cdot \mathbf{a}) = n \cdot f(\mathbf{a}) + 1 \cdot f(\mathbf{a}) = (n+1) \cdot f(\mathbf{a}).$$

∎

Lemma 3.5. *If $f : U \to V$ is additive, then for every rational number $r \in \mathbb{Q}$, $f(r \cdot \mathbf{a}) = r \cdot f(\mathbf{a})$.*

Proof. (The set $\mathbb{Q}$ of rational numbers is the set of fractions with integer numerator and non-zero integer denominator — their decimal expansions are repeating or terminating, so $\mathbb{Q}$ is contained in, but not equal to, the set $\mathbb{R}$.)

Let $r = \frac{p}{q}$ with $p, q \in \mathbb{Z}$, $q > 0$. Using the previous Lemma applied to both p and q,

$$p \cdot f(\mathbf{a}) = f(p \cdot \mathbf{a}) = f((q \cdot \frac{p}{q}) \cdot \mathbf{a}) = f(q \cdot (\frac{p}{q} \cdot \mathbf{a})) = q \cdot f(\frac{p}{q} \cdot \mathbf{a}) \implies \frac{p}{q} \cdot f(\mathbf{a}) = f(\frac{p}{q} \cdot \mathbf{a}).$$

∎

Example 3.6. It is not so hard to construct a function which has the scaling property but which is not linear because it is not additive. Define a function $f : \mathbb{R}^2 \to \mathbb{R}$ by the piecewise formula: $f(x, y) = x$ if $y \neq 0$, and $f(x, y) = 0$ if $y = 0$. Then, to show $f(r \cdot (x, y)) = r \cdot f(x, y)$ for any $r \in \mathbb{R}$, there are three cases to check:

$y \neq 0, r \neq 0 \implies f(r \cdot (x, y)) = f(rx, ry) = rx = r \cdot f(x, y)$.

$y = 0, r \neq 0 \implies f(r \cdot (x, y)) = f(rx, 0) = 0 = r \cdot 0 = r \cdot f(x, y)$.

$r = 0 \implies f(r \cdot (x, y)) = f(0, 0) = 0 = r \cdot f(x, y)$.

However, f is not additive: let $\vec{a} = (1, 1)$ and $\vec{b} = (1, -1)$. Then $f(\vec{a}) = 1$ and $f(\vec{b}) = 1$, so $f(\vec{a}) + f(\vec{b}) = 2$, while $f(\vec{a} + \vec{b}) = f(2, 0) = 0$.

There also exist non-linear functions $f : U \to V$ which are additive but do not have the scaling property for all real scalars; however, these are more difficult to construct. One reason it can get messy is that Lemma 3.5 shows the scaling property must work for all rational scalars, so in such an example, the scaling property could only fail for some irrational scalars.

One can conclude from the Lemma that if we restrict the field of scalars to rationals only, $\mathbb{Q}$, then every additive function is linear. However, for fields of scalars (such as $\mathbb{R}$ or $\mathbb{C}$) that contain but are not equal to $\mathbb{Q}$, there may be an additive function that does not have the scaling property. The following construction gives an example of an additive map $f : \mathbb{R} \to \mathbb{R}$ which does not satisfy $f(r \cdot \vec{a}) = r \cdot f(\vec{a})$ for all $r, \vec{a} \in \mathbb{R}$ (continuing to use the vector notation $\vec{a}$ even though $\vec{a}$ is just a number).

Example 3.7. Step 1. $\mathbb{R}$ is a vector space over the field $\mathbb{Q}$. Use the usual addition on the group $\mathbb{R}$. Also, for any scalar $r \in \mathbb{Q}$, use the usual real number multiplication to define $r \cdot \vec{a} \in \mathbb{R}$. All the vector space axioms are satisfied (Definition 1.1 with $V = \mathbb{R}$, but $\mathbb{R}$ in the scalar multiplication axioms replaced by $\mathbb{Q}$).

Step 2. There exists a basis β for $\mathbb{R}$, as a vector space over $\mathbb{Q}$. Such a basis is called a "Hamel basis" in number theory; the existence requires the Theorem that any vector space over any field has a basis, which is a consequence of the Axiom of Choice from set theory. In particular, any real number is uniquely expressible as a finite sum of rational multiples of elements of β.

Step 3. β is non-empty (since $\mathbb{R}$ is not the trivial space $\{0\}$), and β contains more than 1 element, since $\beta_1 \in \beta \implies \beta_1 \neq 0$ and $\mathrm{span}(\{\beta_1\}) = \{r \cdot \beta_1 : r \in \mathbb{Q}\}$, and this is not all of $\mathbb{R}$ because it does not contain the number $\sqrt{2} \cdot \beta_1$.

Step 4. Let $\beta = \{\beta_1, \beta_2, \ldots\}$ be a basis. To define a function $f : \mathbb{R} \to \mathbb{R}$, we first define the values of f on the basis elements. Define $f(\beta_1) = 1$ and $f(\beta_j) = 0$ for all $j \neq 1$. (I picked this to get a simple example, other than the zero map or the identity map. There could be lots of other choices.) Then, define f for an arbitrary element $\vec{a}$ in $\mathbb{R}$ by expanding $\vec{a}$ as a rational linear combination of basis elements: $\vec{a} = c_1\beta_1 + \ldots c_n\beta_n$, and treating f as if it were additive:

$$f(\vec{a}) = f(c_1\beta_1 + c_2\beta_2 + \ldots + c_n\beta_n) = c_1 f(\beta_1) + c_2 f(\beta_2) \ldots + c_n f(\beta_n) = c_1 \cdot 1 + c_2 \cdot 0 + \ldots + c_n \cdot 0 = c_1.$$

The uniqueness of the coefficients $c_1, \ldots, c_n$ is crucial, for f to be a well-defined function.

Step 5. Then it is easy to check that f really is additive: if $\vec{a} = c_1\beta_1 + \ldots c_n\beta_n$ and $\vec{b} = b_1\beta_1 + \ldots + b_N\beta_N$, then $\vec{a} + \vec{b} = (c_1\beta_1 + \ldots c_n\beta_n) + (b_1\beta_1 + \ldots + b_N\beta_N) = (c_1 + b_1)\beta_1 + \ldots$, and $f(\vec{a} + \vec{b}) = c_1 + b_1 = f(\vec{a}) + f(\vec{b})$.

Step 6. The above steps define a map $f : \mathbb{R} \to \mathbb{R}$. It actually is linear when $\mathbb{R}$ is considered as a vector space over $\mathbb{Q}$ (allowing only rational scalar multiples), but it is not linear when $\mathbb{R}$ is considered as a vector space over $\mathbb{R}$. That is, it does not have the scaling property for all real scalars, as the following example shows. Consider $r = \sqrt{2} \in \mathbb{R}$ and $\vec{a} = \beta_1 \in \mathbb{R}$. Since $\sqrt{2}\beta_1$ is a real number, it is equal to $c_1\beta_1 + \ldots + c_n\beta_n$ for some rational coefficients $c_1, \ldots, c_n$. Then $r \cdot f(\vec{a}) = \sqrt{2} \cdot f(1 \cdot \beta_1) = \sqrt{2} \cdot 1 = \sqrt{2}$, but $f(r \cdot \vec{a}) = f(\sqrt{2}\beta_1) = f(c_1\beta_1 + \ldots + c_n\beta_n) = c_1$. Since $c_1 \in \mathbb{Q}$ and $\sqrt{2} \notin \mathbb{Q}$, f is not linear.

Remark: The above Step 2 does not actually construct the basis β, it merely asserts its existence. So, the definition of f doesn't say explicitly what number would result from plugging in a specific number x into f, for example: $f(\sqrt{5} + 2\pi) =???$

Exercise 3.8. Considering $\mathbb{R}$ as a vector space over the field $\mathbb{Q}$, show that $\sqrt{3} \notin \mathrm{span}(\{1, \sqrt{2}\})$.

Hint: you may assume that $\sqrt{3}$, $\sqrt{2} \notin \mathbb{Q}$, and $\sqrt{3} \notin \mathrm{span}(\{\sqrt{2}\})$.

4 Distance functions

The following notion of distance measurement applies to any set (like a sphere, plane, or vector space).

Definition 4.1. A "distance function" on a set P is a function $d : P \times P \to \mathbb{R}$ such that

- $d(x, y) \geq 0$
- $x = y \implies d(x, y) = 0$.
- $d(x, y) = 0 \implies x = y$.
- $d(x, y) = d(y, x)$.
- $d(x, z) \leq d(x, y) + d(y, z)$.

Example 4.2. Let $P = \mathbb{R}^n$; then for column vectors $\vec{x} = (x_1, x_2, \ldots, x_n)_{n\times 1}$, $\vec{y} = (y_1, y_2, \ldots, y_n)_{n\times 1}$, the following is a distance function:

$$d(\vec{x}, \vec{y}) = \sqrt{(y_1 - x_1)^2 + \ldots + (y_n - x_n)^2}$$

$$= \sqrt{(\vec{y} - \vec{x}) \cdot (\vec{y} - \vec{x})} = \sqrt{(\vec{y} - \vec{x})^T(\vec{y} - \vec{x})},$$

where $\vec{a} \cdot \vec{b}$ is the "dot product," equal to the matrix multiplication of the row $(\vec{a})^T$ and the column $\vec{b}$.

This particular distance function in the Example is not unique, but its simple formula makes it convenient, and it is the same as the "Euclidean" distance familiar from pre-calculus in $\mathbb{R}^1$ and $\mathbb{R}^2$, and from multivariable calculus in $\mathbb{R}^3$. One way in which it is not unique, for example, is that multiplying the above function d by any positive constant gives another function which still satisfies all five properties. This corresponds to a "choice of scale" of the vector space P, and our choice is that the vector $(1, 0, 0, \ldots, 0)$ has length 1.

Exercise 4.3. Given any set P and any distance function d, suppose $\alpha : P \to P$ is a function such that $d(\alpha(x), \alpha(y)) = d(x, y)$ for all $x, y \in P$. Show α must be one-to-one. Give an example of a set P and a function α which satisfies the equality, but which is not "onto."

Definition 4.4. A function $\alpha : P \to P$ such that α is onto, and $d(\alpha(x), \alpha(y)) = d(x, y)$ for all $x, y \in P$, is called a "motion of P." The set of all such functions is denoted $M(P, d)$.

Exercise 4.5. Any motion of P must be an invertible function (why?), and the inverse function is also a motion of P. If $\alpha : P \to P$ and $\beta : P \to P$ are motions of P, then so is the composite function $\beta \circ \alpha : P \to P$.

Definition 4.6. Given a real vector space V, a function $n : V \to \mathbb{R}$ is a "norm" means: n has the following properties:

- $n(\mathbf{v}) \geq 0$
- $n(k \cdot \mathbf{v}) = |k|n(\mathbf{v})$.
- $n(\mathbf{u} + \mathbf{v}) \leq n(\mathbf{u}) + n(\mathbf{v})$.
- $n(\mathbf{v}) = 0 \implies \mathbf{v} = \mathbf{0}$.

Proposition 4.7. *Given a real vector space V and a norm n, the function*

$$d(\mathbf{u}, \mathbf{v}) = n(\mathbf{v} - \mathbf{u})$$

is a distance function on V. ∎

It is often convenient to allow complex numbers in linear algebra. Let $\mathbb{C}$ denote the set of all complex numbers, $\mathbb{C} = \{x + iy : x, y \in \mathbb{R}, i^2 = -1\}$, which includes the set of real numbers $\mathbb{R}$ as a subset (numbers with $y = 0$). The "complex conjugation" function $\mathbb{C} \to \mathbb{C}$ is denoted by a "bar," so if $z = x + iy$, then $\bar{z} = \overline{x + iy} = x - iy$. Complex conjugation satisfies the following identities for $z, w \in \mathbb{C}$: $\overline{z + w} = \bar{z} + \bar{w}$, $\overline{zw} = \bar{z}\bar{w}$.

Definition 1.1, of a "real vector space," can be modified to define a "complex vector space," in which scalar multiplication allows complex numbers as scalars. The first four properties listed in Definition 1.1, on addition, are the same, and the remaining properties involving scalar multiplication can be modified only by changing the set of scalars $\mathbb{R}$ to the new set of complex scalars, $\mathbb{C}$.

Definition 3.2 can also be modified: where U and V are both complex vector spaces, an additive map $f : U \to V$ is "linear over $\mathbb{C}$" means: $f(z \cdot \mathbf{a}) = z \cdot f(\mathbf{a})$ for all $z \in \mathbb{C}$.

Let $\mathbb{C}^n$ denote the set of vectors with n complex components — it is an example of a complex vector space.

In later statements which could apply to either $\mathbb{R}^n$ or $\mathbb{C}^n$, I'll use $\mathbb{K}^n$ to denote either the real vector space $\mathbb{R}^n$ or the complex vector space $\mathbb{C}^n$, and $\mathbb{K}$ for the set of scalars (either $\mathbb{R}$ or $\mathbb{C}$).

The definition of "norm" can also be adapted to complex vector spaces; the input is an element of a complex vector space, but the output is still a nonnegative real number. The only modification to Definition 4.6 is that the "absolute value" refers to the complex number version of absolute value: if $z = x + iy$, then $|z|$ is defined by the formula $|z| = \sqrt{x^2 + y^2}$ (the same quantity appearing as the radius in the polar coordinate formula). Proposition 4.7 also can be applied to a complex vector space V, so that a norm on V defines a distance function on the set V using the same formula.

5 Bilinear forms and sesquilinear forms

The "dot product" in $\mathbb{R}^n$ was useful for understanding the connections between the geometry and the algebra of vectors. It is called a "product" because it takes two vectors and calculates a scalar. This idea can be generalized by considering other functions which take two vectors as input and give one scalar as output.

The notation for the dot product, $\vec{x} \cdot \vec{y}$, could be replaced by the "bracket" notation, $\langle \vec{x}, \vec{y} \rangle$. However, to emphasize that we are working with a multivariable function that takes two input vectors $\vec{x} \in V$, $\vec{y} \in V$, and gives scalar output (in $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$), we call the function "g," and use the function notation $g : V \times V \to \mathbb{K}$, and the expression $\langle \vec{x}, \vec{y} \rangle = g(\vec{x}, \vec{y})$.

Definition 5.1. A function $g : V \times V \to \mathbb{K}$ is "bi-additive" means: g satisfies both identities:

- $g(\mathbf{x}, \mathbf{y} + \mathbf{z}) = g(\mathbf{x}, \mathbf{y}) + g(\mathbf{x}, \mathbf{z})$,
- $g(\mathbf{x} + \mathbf{z}, \mathbf{y}) = g(\mathbf{x}, \mathbf{y}) + g(\mathbf{z}, \mathbf{y})$.

Definition 5.2. A function $g : V \times V \to \mathbb{K}$ is "bilinear" means: g is bi-additive, and g also satisfies these two identities for any $k \in \mathbb{K}$:

- $g(k \cdot \mathbf{x}, \mathbf{y}) = k \cdot g(\mathbf{x}, \mathbf{y})$,
- $g(\mathbf{x}, k \cdot \mathbf{y}) = k \cdot g(\mathbf{x}, \mathbf{y})$.

A function which is bilinear is also called a "bilinear form."

Exercise 5.3. Given a bi-additive function g on a vector space V, for all vectors $\mathbf{x} \in V$, $g(\mathbf{x}, \mathbf{0}) = g(\mathbf{0}, \mathbf{x}) = 0$. Also, $g(k \cdot \mathbf{x}, \mathbf{y}) = g(\mathbf{x}, k \cdot \mathbf{y}) = k \cdot g(\mathbf{x}, \mathbf{y})$ for all rational scalars $k \in \mathbb{Q}$. It further follows that $g(\mathbf{x}, \mathbf{y} - \mathbf{z}) = g(\mathbf{x}, \mathbf{y}) - g(\mathbf{x}, \mathbf{z})$ and $g(\mathbf{x} - \mathbf{z}, \mathbf{y}) = g(\mathbf{x}, \mathbf{y}) - g(\mathbf{z}, \mathbf{y})$.

Exercise 5.4. Any bi-additive function g satisfies the following identity:

$$\begin{aligned} & g(\mathbf{u} + \mathbf{v} - \mathbf{w}, \mathbf{u} + \mathbf{v} - \mathbf{w}) \\ = \; & g(\mathbf{u} - \mathbf{w}, \mathbf{u} - \mathbf{w}) + g(\mathbf{v} - \mathbf{w}, \mathbf{v} - \mathbf{w}) - g(\mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v}) + g(\mathbf{u}, \mathbf{u}) + g(\mathbf{v}, \mathbf{v}) - g(\mathbf{w}, \mathbf{w}). \end{aligned}$$

Exercise 5.5. Any bilinear function g satisfies the following identity, for vectors $\mathbf{u}, \mathbf{w} \in V$ and scalars $\lambda \in \mathbb{K}$:

$$g(\lambda \cdot \mathbf{u} - \mathbf{w}, \lambda \cdot \mathbf{u} - \mathbf{w}) = (1 - \lambda)(g(\mathbf{w}, \mathbf{w}) - \lambda g(\mathbf{u}, \mathbf{u})) + \lambda g(\mathbf{u} - \mathbf{w}, \mathbf{u} - \mathbf{w}).$$

The following Theorem deals with the special case that the vector space V is $\mathbb{K}^n$, so its elements are column vectors. Then any bilinear form can be expressed as a certain kind of matrix product.

Theorem 5.6. *If g is a bilinear form on $\mathbb{K}^n$, then there exists a matrix $G_{n \times n}$ with entries in $\mathbb{K}$ such that*

$$g(\mathbf{x}, \mathbf{y}) = \mathbf{y}^T_{1 \times n} G_{n \times n} \mathbf{x}_{n \times 1}.$$

Proof. Recall the standard basis of column vectors $\{\mathbf{e}^\ell = (0,\dots,0,1,0,\dots,0)_{n\times 1}\}$, for $\ell = 1,\dots,n$. Define the entries of G by the formula $G_{j\ell} = g(\mathbf{e}^\ell, \mathbf{e}^j)$. For column vectors $\mathbf{x} = \sum x_\ell \mathbf{e}^\ell$ and $\mathbf{y} = \sum y_j \mathbf{e}^j$, the bilinearity properties give

$$g(\mathbf{x},\mathbf{y}) = \sum_{j=1}^{n} y_j \left(\sum_{\ell=1}^{n} x_\ell g(\mathbf{e}^\ell, \mathbf{e}^j) \right) = \sum_{j=1}^{n} y_j \left(\sum_{\ell=1}^{n} G_{j\ell} x_\ell \right) = \mathbf{y}^T G \mathbf{x},$$

where $\mathbf{y}^T$ is a row vector (the transpose of the column vector $\mathbf{y}$). ∎

Theorem 5.7. *Given a bi-additive function g on a vector space V, the following are equivalent:*

1. *For all vectors $\mathbf{x} \in V$, $g(\mathbf{x},\mathbf{x}) = 0$;*
2. *For all $\mathbf{x}$, $\mathbf{y} \in V$, $g(\mathbf{x},\mathbf{y}) = -g(\mathbf{y},\mathbf{x})$.*

Proof. To show *2.* $\implies$ *1.*, just plug $\mathbf{y} = \mathbf{x}$ into *2.* to get $g(\mathbf{x},\mathbf{x}) = -g(\mathbf{x},\mathbf{x})$, which implies *1.*

For *1.* $\implies$ *2.*, expand using the bi-additive property:

$$\begin{aligned} g(\mathbf{x}+\mathbf{y},\mathbf{x}+\mathbf{y}) &= g(\mathbf{x},\mathbf{x}) + g(\mathbf{y},\mathbf{x}) + g(\mathbf{x},\mathbf{y}) + g(\mathbf{y},\mathbf{y}) \\ \implies 0 &= 0 + g(\mathbf{y},\mathbf{x}) + g(\mathbf{x},\mathbf{y}) + 0, \end{aligned}$$

and *2.* follows. ∎

Definition 5.8. A function $g : V \times V \to \mathbb{K}$ which is bi-additive and satisfies $g(\mathbf{x},\mathbf{x}) = 0$ for all $\mathbf{x} \in V$, and also $g(\lambda \cdot \mathbf{y}, \mathbf{x}) = \lambda g(\mathbf{y},\mathbf{x})$ for all $\mathbf{x}$, $\mathbf{y} \in V$, $\lambda \in \mathbb{K}$, is called "alternating."

Theorem 5.9. *If $g : V \times V \to \mathbb{K}$ is alternating, then it is a bilinear form.*

Proof. Applying Theorem 5.7 to the bi-additive function g, for all $\mathbf{x}$, $\mathbf{y} \in V$, $g(\mathbf{x},\mathbf{y}) = -g(\mathbf{y},\mathbf{x})$. So, for any scalar $\lambda \in \mathbb{K}$, $g(\mathbf{x}, \lambda\cdot\mathbf{y}) = -g(\lambda\cdot\mathbf{y},\mathbf{x}) = -(\lambda g(\mathbf{y},\mathbf{x})) = -(\lambda(-g(\mathbf{x},\mathbf{y}))) = \lambda g(\mathbf{x},\mathbf{y})$. ∎

So, a bi-additive function which is alternating can be called an alternating bilinear form or an "alternating form."

Theorem 5.10. *Given a complex vector space V and a bi-additive function $g : V \times V \to \mathbb{C}$ satisfying $g(\lambda \cdot \mathbf{y}, \mathbf{x}) = \lambda g(\mathbf{y},\mathbf{x})$ for all $\mathbf{x}$, $\mathbf{y} \in V$, $\lambda \in \mathbb{C}$, the following are equivalent:*

1. *For all vectors $\mathbf{x} \in V$, $g(\mathbf{x},\mathbf{x}) = 0$;*
2. *For all $\mathbf{x}$, $\mathbf{y} \in V$, $g(\mathbf{x},\mathbf{y}) = -g(\mathbf{y},\mathbf{x})$;*
3. *g is an alternating form;*
4. *g is a bilinear form and $g(\mathbf{x},\mathbf{x})$ is real for all $\mathbf{x} \in V$.*

Proof. The equivalence *1.* $\iff$ *2.* was established in Theorem 5.7, and *1.* $\iff$ *3.* is Definition 5.8. For the implication *1.* $\implies$ *4.*, if $g(\mathbf{x},\mathbf{x}) = 0$, then obviously $g(\mathbf{x},\mathbf{x})$ is real, and the bilinear property was proved in Theorem 5.9.

To show *4.* $\implies$ *1.*, consider any $\mathbf{x} \in V$, so $g(\mathbf{x},\mathbf{x})$ is real. The number $g((1+i)\cdot\mathbf{x}, (1+i)\cdot\mathbf{x})$ is also real, and using the bilinearity property:

$$g((1+i)\cdot\mathbf{x}, (1+i)\cdot\mathbf{x}) = (1+i)^2 g(\mathbf{x},\mathbf{x}) = 2ig(\mathbf{x},\mathbf{x}),$$

so we can conclude $ig(\mathbf{x},\mathbf{x})$ is real. However, the only complex number which is real and whose product with i is also real is 0, so $g(\mathbf{x},\mathbf{x}) = 0$. ∎

Definition 5.11. A function $g : V \times V \to \mathbb{K}$ is "sesquilinear" means: g is bi-additive, and g also satisfies these two identities for any $k \in \mathbb{K}$:

- $g(k \cdot \mathbf{x}, \mathbf{y}) = k \cdot g(\mathbf{x},\mathbf{y})$,
- $g(\mathbf{x}, k \cdot \mathbf{y}) = \overline{k} \cdot g(\mathbf{x},\mathbf{y})$.

Note that the first identity is the same as in Definition 5.2, and the second one involves the complex conjugate of the scalar k on the RHS. A function which is sesquilinear is also called a "sesquilinear form."

If $\mathbb{K} = \mathbb{R}$, that is, in the case where V is a real vector space and g is a real-valued function, then only real scalars k are allowed, and the complex conjugate of the real scalar k is equal to k itself ($\overline{k} = \overline{k + 0i} = k - 0i = k$). So in the real case, bilinear and sesquilinear mean the same thing. In either the real or complex case, the Definition is consistent with the result from Exercise 5.3 that $g(\mathbf{x}, k \cdot \mathbf{y}) = k \cdot g(\mathbf{x},\mathbf{y})$ for rational k, since all rational numbers are real and satisfy $\overline{k} = k$.

Theorem 5.12. *Given a complex vector space V, and a function $g : V \times V \to \mathbb{C}$, the following are equivalent:*

1. *g is sesquilinear and for all vectors $\mathbf{x} \in V$, $g(\mathbf{x},\mathbf{x}) = 0$;*
2. *g is sesquilinear and for all $\mathbf{x}$, $\mathbf{y} \in V$, $g(\mathbf{x},\mathbf{y}) = -g(\mathbf{y},\mathbf{x})$;*
3. *g is both alternating and sesquilinear;*
4. *g is both bilinear and sesquilinear;*
5. *g is the constant function zero.*

Proof. Since sesquilinear functions are bi-additive, *1.* $\iff$ *2.* by Theorem 5.7. Since sesquilinear functions also satisfy $g(k \cdot \mathbf{x}, \mathbf{y}) = k \cdot g(\mathbf{x},\mathbf{y})$, *1.* $\iff$ *3.* by Definition 5.8, and also Theorem 5.9 applies, so g is bilinear and *1.* $\implies$ *4.*

To show *4.* $\implies$ *5.*, for any $\mathbf{x}$, $\mathbf{y} \in V$, $g(\mathbf{x}, i \cdot \mathbf{y}) = ig(\mathbf{x},\mathbf{y})$ by the bilinear property, but $g(\mathbf{x}, i \cdot \mathbf{y}) = (-i)g(\mathbf{x},\mathbf{y})$ by the sesquilinear property. Dividing by i gives $g(\mathbf{x},\mathbf{y}) = -g(\mathbf{x},\mathbf{y})$, so $g(\mathbf{x},\mathbf{y}) = 0$.

Finally, *5.* $\implies$ *1.* is obvious. ∎

In analogy with Theorem 5.6, any sesquilinear form on the vector space $\mathbb{K}^n$ can be represented by a combination of complex conjugation and matrix multiplication.

Theorem 5.13. *If g is a sesquilinear form on $\mathbb{K}^n$, then there exists a matrix $G_{n\times n}$ with entries in $\mathbb{K}$ such that*

$$g(\mathbf{x},\mathbf{y}) = \overline{\mathbf{y}}^T_{1\times n} G_{n\times n} \mathbf{x}.$$

Proof. Use the same standard basis vectors as the previous Theorem: $\{\mathbf{e}^\ell = (0, \ldots, 0, 1, 0, \ldots, 0)\}$, for $\ell = 1, \ldots, n$. Define the entries of G by the formula $G_{j\ell} = g(\mathbf{e}^\ell, \mathbf{e}^j)$. For $\mathbf{x} = \sum x_\ell \mathbf{e}^\ell$ and $\mathbf{y} = \sum y_j \mathbf{e}^j$, the sesquilinear properties give

$$g(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^{n} \overline{y_j} \left(\sum_{\ell=1}^{n} x_\ell g(\mathbf{e}^\ell, \mathbf{e}^j) \right) = \sum_{j=1}^{n} \overline{y_j} \left(\sum_{\ell=1}^{n} G_{j\ell} x_\ell \right) = \overline{\mathbf{y}}^T G \mathbf{x},$$

where $\overline{\mathbf{y}}^T$ denotes the row vector listing the complex conjugates of the entries of $\mathbf{y}$: $(\overline{y_1}, \ldots, \overline{y_n})$. ∎

In the real case, this gives the same real matrix G as Theorem 5.6.

Here are three properties which a bilinear form or a sesquilinear form $g : V \times V \to \mathbb{K}$ could have:

Definition 5.14. g is "non-degenerate" means: for each non-zero vector $\mathbf{x} \in V$, $\mathbf{x} \neq \mathbf{0}$, there exists a vector $\mathbf{y} \in V$ so that $g(\mathbf{x}, \mathbf{y}) \neq 0$.

Definition 5.15. g is "positive semidefinite" means: for all $\mathbf{x} \in V$, $g(\mathbf{x}, \mathbf{x}) \geq 0$.

Definition 5.16. g is "positive definite" means: for every non-zero $\mathbf{x} \in V$, $g(\mathbf{x}, \mathbf{x}) > 0$.

Example 5.17. The function which always gives output 0 ($g(\mathbf{x}, \mathbf{y}) = 0$ for all $\mathbf{x}, \mathbf{y} \in V$) is both bilinear and sesquilinear (as in Theorem 5.12), and it is positive semidefinite. However, unless V is a zero-dimensional vector space, the zero function is neither positive definite nor non-degenerate.

The following Theorem applies to the case where the vector space V is $\mathbb{K}^n$.

Theorem 5.18. *Given a bilinear or sesquilinear form g on $\mathbb{K}^n$, let G be the matrix from the corresponding Theorem (5.6 or 5.13). g is non-degenerate if and only if G is invertible.*

Proof. G is not invertible if and only if there is some element $\mathbf{x} \neq \mathbf{0}$ in the nullspace of G. In the sesquilinear case, for any $\mathbf{y} \in \mathbb{K}^n$,

$$g(\mathbf{x}, \mathbf{y}) = \overline{\mathbf{y}}^T G \mathbf{x} = \overline{\mathbf{y}}^T \mathbf{0} = 0,$$

which is equivalent to g not being non-degenerate. The equation for the bilinear case is similar (delete the bar). ∎

Theorem 5.19. *For any real or complex vector space V, if a bilinear or sesquilinear form g is positive definite, then it is non-degenerate and positive semidefinite.*

Proof. The non-degeneracy follows immediately from the definitions — just choose $\mathbf{y} = \mathbf{x}$ in the definition of non-degenerate. For positive semidefinite, it remains only to check $g(\mathbf{0}, \mathbf{0}) = 0$, which follows from Exercise 5.3. ∎

Example 5.20. The converse of the above Theorem does not hold; a bilinear form can be non-degenerate and positive semidefinite without being positive definite. For $V = \mathbb{R}^2$, and $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $\vec{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$, define

$$g(\vec{x}, \vec{y}) = x_1 y_1 + 2x_2 y_1 + x_2 y_2 = (y_1 \;\; y_2)_{1\times 2} \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}_{2\times 2} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_{2\times 1}.$$

It is easy to check that g is bilinear, and since the coefficient matrix is non-singular, g is non-degenerate by Theorem 5.18. Further, g is positive semidefinite, since

$$g(\vec{x},\vec{x}) = x_1^2 + 2x_1x_2 + x_2^2 = (x_1+x_2)^2 \geq 0,$$

but not positive definite, since if $\vec{x} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, then $g(\vec{x},\vec{x}) = 0$.

Theorem 5.21. *Given a non-degenerate bilinear or sesquilinear form g on V, and vectors $\mathbf{x}$, $\mathbf{z} \in V$, if $g(\mathbf{x},\mathbf{y}) = g(\mathbf{z},\mathbf{y})$ for all $\mathbf{y} \in V$, then $\mathbf{x} = \mathbf{z}$.*

Proof. If $g(\mathbf{x},\mathbf{y}) = g(\mathbf{z},\mathbf{y})$ for all $\mathbf{y} \in V$, then $g(\mathbf{x}-\mathbf{z},\mathbf{y}) = g(\mathbf{x},\mathbf{y}) - g(\mathbf{z},\mathbf{y}) = 0$ for all $\mathbf{y} \in V$ Since $\mathbf{x}-\mathbf{z} \neq \mathbf{0}$ would contradict the definition of non-degenerate, we can conclude $\mathbf{x}-\mathbf{z} = \mathbf{0}$. ∎

Theorem 5.22. *Given a non-degenerate bilinear or sesquilinear form g on V, suppose $H : V \to V$ is any function which is "onto" (for any $\mathbf{y} \in V$, there exists $\mathbf{w} \in V$ such that $H(\mathbf{w}) = \mathbf{y}$), and which satisfies the equation*

$$g(H(\mathbf{x}), H(\mathbf{y})) = g(\mathbf{x},\mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in V$. Then H is a linear function.

Proof. There are two parts from the definition of linear to check. Suppose $\mathbf{x} \in V$ and $\lambda \in \mathbb{K}$. Then, for any $\mathbf{y} \in V$, there exists $\mathbf{w} \in V$ such that $H(\mathbf{w}) = \mathbf{y}$, and

$$g(H(\lambda\cdot\mathbf{x}),\mathbf{y}) = g(H(\lambda\cdot\mathbf{x}),H(\mathbf{w})) = g(\lambda\cdot\mathbf{x},\mathbf{w}) = \lambda g(\mathbf{x},\mathbf{w}) = \lambda g(H(\mathbf{x}),H(\mathbf{w})) = g(\lambda\cdot H(\mathbf{x}),\mathbf{y}).$$

By the previous Theorem, this shows $H(\lambda\cdot\mathbf{x}) = \lambda\cdot H(\mathbf{x})$.

Given $\mathbf{x},\mathbf{z} \in V$, and $H(\mathbf{w}) = \mathbf{y}$ as above,

$$\begin{aligned} g(H(\mathbf{x}+\mathbf{z}),\mathbf{y}) &= g(H(\mathbf{x}+\mathbf{z}),H(\mathbf{w})) = g(\mathbf{x}+\mathbf{z},\mathbf{w}) \\ &= g(\mathbf{x},\mathbf{w}) + g(\mathbf{z},\mathbf{w}) = g(H(\mathbf{x}),H(\mathbf{w})) + g(H(\mathbf{z}),H(\mathbf{w})) \\ &= g(H(\mathbf{x})+H(\mathbf{z}),H(\mathbf{w})) = g(H(\mathbf{x})+H(\mathbf{z}),\mathbf{y}), \end{aligned}$$

and again by Theorem 5.21, we can conclude $H(\mathbf{x}+\mathbf{z}) = H(\mathbf{x}) + H(\mathbf{z})$, so H is linear. ∎

Definition 5.23. Given a function $g : V \times V \to \mathbb{K}$, define a function $q : V \to \mathbb{K}$ by the formula:

$$q(\mathbf{x}) = g(\mathbf{x},\mathbf{x}).$$

q is a "quadratic form" means that $q(\mathbf{x}) = g(\mathbf{x},\mathbf{x})$ for some bilinear or sesquilinear form g.

In general, a quadratic form is not linear, but does satisfy $q(\mathbf{0}) = 0$ by Exercise 5.3. The expression $g(\mathbf{x},\mathbf{x})$ appeared already in Definitions 5.15, 5.16.

Notation 5.24. Given a function $g : V \times V \to \mathbb{K}$, define a function $\psi : V \times V \to \mathbb{K}$ by the formula:

$$\psi(\mathbf{x},\mathbf{y}) = g(\mathbf{y}-\mathbf{x},\mathbf{y}-\mathbf{x}).$$

The functions q and ψ defined by g are related by the identity $\psi(\mathbf{0},\mathbf{y}) = g(\mathbf{y},\mathbf{y}) = q(\mathbf{y})$, and also if g is bi-additive, then:

$$\begin{aligned}\psi(\mathbf{x},\mathbf{y}) &= q(\mathbf{y}-\mathbf{x}) = g(\mathbf{y}-\mathbf{x},\mathbf{y}-\mathbf{x}) \\ &= g(\mathbf{y},\mathbf{y}) - g(\mathbf{y},\mathbf{x}) - g(\mathbf{x},\mathbf{y}) + g(\mathbf{x},\mathbf{x}) \\ &= q(\mathbf{y}) - g(\mathbf{y},\mathbf{x}) - g(\mathbf{x},\mathbf{y}) + q(\mathbf{x}).\end{aligned} \tag{1}$$

Theorem 5.25. *Given a bi-additive function g on a vector space V, let q and ψ be defined as above. If $H : V \to V$ is any function satisfying $g(H(\mathbf{x}),H(\mathbf{y})) = g(\mathbf{x},\mathbf{y})$ for all $\mathbf{x},\mathbf{y} \in V$, then H also satisfies $q(H(\mathbf{x})) = q(\mathbf{x})$ and $\psi(H(\mathbf{x}),H(\mathbf{y})) = \psi(\mathbf{x},\mathbf{y})$.*

Proof. The first claim follows immediately from the hypothesis and the definition of q:

$$q(H(\mathbf{x})) = g(H(\mathbf{x}),H(\mathbf{x})) = g(\mathbf{x},\mathbf{x}) = q(\mathbf{x}).$$

The second claim follows from the above Equation (1):

$$\begin{aligned}\psi(H(\mathbf{x}),H(\mathbf{y})) &= g(H(\mathbf{y}),H(\mathbf{y})) - g(H(\mathbf{y}),H(\mathbf{x})) - g(H(\mathbf{x}),H(\mathbf{y})) + g(H(\mathbf{x}),H(\mathbf{x})) \\ &= g(\mathbf{y},\mathbf{y}) - g(\mathbf{y},\mathbf{x}) - g(\mathbf{x},\mathbf{y}) + g(\mathbf{x},\mathbf{x}) \\ &= \psi(\mathbf{x},\mathbf{y}).\end{aligned}$$

∎

Notation 5.26. Every complex number $z \in \mathbb{C}$ is, by definition, of the form $z = x + iy$, where x and y are real numbers. The function that takes input $z \in \mathbb{C}$ and gives output x is called the "real part" function, denoted $\mathrm{Re}(z) = x$. Similarly, the "imaginary part" function is denoted $\mathrm{Im}(z) = y$.

Notation 5.27. For a vector $\mathbf{v}$ in a complex vector space V, scalar multiplication is defined for any complex number scalar z, and denoted by $z \cdot \mathbf{v}$. If we agree to forget about complex numbers and work only with real scalars ($z = x + iy$ is real exactly when $y = 0$, so $z = x + i0 = x$), then the same set of vectors V still satisfies the axioms defining a real vector space. It makes sense to refer to "real linear" functions or "real bilinear functions" when they are linear or bilinear only for real scalars (but not necessarily for all complex scalars).

Example 5.28. If V is a complex vector space, and $g : V \times V \to \mathbb{C}$ is a complex bilinear or sesquilinear form, then the following composite function is defined: $\mathrm{Re} \circ g : V \times V \to \mathbb{R}$. This composition has formula $(\mathrm{Re} \circ g)(\mathbf{u},\mathbf{v}) = \mathrm{Re}(g(\mathbf{u},\mathbf{v}))$, so it takes two vectors in V as input and returns a real scalar as output. The function $\mathrm{Re} \circ g$ is a real bilinear form, on the set V considered as a real vector space: it is bi-additive, and for real scalars r, $(\mathrm{Re} \circ g)(r \cdot \mathbf{u},\mathbf{v}) = (\mathrm{Re} \circ g)(\mathbf{u}, r \cdot \mathbf{v}) = r((\mathrm{Re} \circ g)(\mathbf{u},\mathbf{v}))$.

Example 5.29. If V is a complex vector space, and $g : V \times V \to \mathbb{C}$ is a complex bilinear or sesquilinear form, which is positive definite (or semidefinite), then the real bilinear form $\mathrm{Re} \circ g$ from the previous Example is positive definite (or semidefinite): since positive numbers are real numbers, if $g(\mathbf{u},\mathbf{u}) > 0$, then $(\mathrm{Re} \circ g)(\mathbf{u},\mathbf{u}) = \mathrm{Re}(g(\mathbf{u},\mathbf{u})) = g(\mathbf{u},\mathbf{u}) > 0$.

6 Inner Products

Definition 6.1. An "inner product" on a vector space V is a function g which takes two input vectors (an ordered pair $(\mathbf{x}, \mathbf{y})$, with $\mathbf{x}$ and $\mathbf{y} \in V$), and which gives just one number as output, satisfying three properties:

- For all vectors $\mathbf{x}$, $\mathbf{y} \in V$, and scalars $\lambda \in \mathbb{K}$, $g(\lambda \cdot \mathbf{x}, \mathbf{y}) = \lambda g(\mathbf{x}, \mathbf{y})$.
- For all $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z} \in V$, $g(\mathbf{x} + \mathbf{y}, \mathbf{z}) = g(\mathbf{x}, \mathbf{z}) + g(\mathbf{y}, \mathbf{z})$.
- The function g also must have one of the following two properties:
  - ⋄ (Symmetric) For all $\mathbf{x}$, $\mathbf{y} \in V$, $g(\mathbf{x}, \mathbf{y}) = g(\mathbf{y}, \mathbf{x})$,
  - ⋄ (Hermitian) For all $\mathbf{x}$, $\mathbf{y} \in V$, $g(\mathbf{x}, \mathbf{y}) = \overline{g(\mathbf{y}, \mathbf{x})}$. (switching the order of the inputs gives the complex conjugate output.)

Note that the first two properties are part (but not all) of the definitions of "bilinear" and "sesquilinear."

There are three cases of inner products:

1. Complex Hermitian inner product: the input of g is two vectors from a complex vector space, the output is a complex number, and g has the Hermitian property.
2. Complex symmetric inner product: the input of g is two vectors from a complex vector space, the output is a complex number, and g has the symmetric property.
3. Real inner product: the input of g is two vectors from a real vector space (so the scalar λ from the first property has to be real), and the output is a real number. The two choices of symmetry property are actually identical — g is both symmetric and Hermitian, since every real number is equal to its complex conjugate.

Proposition 6.2. *For an inner product g on a real or complex vector space V,*

- *For all vectors $\mathbf{x} \in V$, $g(\mathbf{x}, \mathbf{0}) = g(\mathbf{0}, \mathbf{x}) = 0$.*
- *For all $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z} \in V$, $g(\mathbf{x}, \mathbf{y} + \mathbf{z}) = g(\mathbf{x}, \mathbf{y}) + g(\mathbf{x}, \mathbf{z})$.*
- *If g is symmetric, then for all vectors $\mathbf{x}$, $\mathbf{y} \in V$, and scalars $\lambda \in \mathbb{K}$, $g(\mathbf{x}, \lambda \cdot \mathbf{y}) = \lambda g(\mathbf{x}, \mathbf{y})$.*

$\blacksquare$

So, all inner products are bi-additive, and a symmetric inner product is always a bilinear form.

Theorem 6.3. *Given a bilinear form $g : V \times V \to \mathbb{K}$, the following are equivalent:*

1. *For all $\mathbf{x}, \mathbf{y} \in V$, if $g(\mathbf{x}, \mathbf{y}) = 0$, then $g(\mathbf{y}, \mathbf{x}) = 0$;*
2. *g is either a symmetric inner product ($g(\mathbf{x}, \mathbf{y}) = g(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x}, \mathbf{y}$), or g is an alternating form ($g(\mathbf{x}, \mathbf{y}) = -g(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x}, \mathbf{y}$).*

Proof. The proof of *2.* $\Longrightarrow$ *1.* is left as an easy exercise.

For *1.* $\Longrightarrow$ *2.*, this proof is based on [J] §6.1. A bilinear form g satisfies the following identity for any $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$:

$$g(\mathbf{x}, g(\mathbf{x}, \mathbf{y}) \cdot \mathbf{z} - g(\mathbf{x}, \mathbf{z}) \cdot \mathbf{y}) = g(\mathbf{x}, \mathbf{y})g(\mathbf{x}, \mathbf{z}) - g(\mathbf{x}, \mathbf{z})g(\mathbf{x}, \mathbf{y}) = 0,$$

so by *1.*, these identities hold:

$$\begin{aligned} g(g(\mathbf{x}, \mathbf{y}) \cdot \mathbf{z} - g(\mathbf{x}, \mathbf{z}) \cdot \mathbf{y}, \mathbf{x}) &= 0 \\ \Longrightarrow\ g(\mathbf{x}, \mathbf{y})g(\mathbf{z}, \mathbf{x}) - g(\mathbf{x}, \mathbf{z})g(\mathbf{y}, \mathbf{x}) &= 0. \end{aligned} \tag{2}$$

In the case $\mathbf{y} = \mathbf{x}$, we can factor to get this identity for all $\mathbf{x}, \mathbf{z} \in V$:

$$(g(\mathbf{z}, \mathbf{x}) - g(\mathbf{x}, \mathbf{z}))g(\mathbf{x}, \mathbf{x}) = 0. \tag{3}$$

Suppose, toward a contradiction, that g is neither symmetric nor alternating. Then there exist $\mathbf{u}, \mathbf{v} \in V$ so that $g(\mathbf{u}, \mathbf{v}) \neq g(\mathbf{v}, \mathbf{u})$, and (by Theorem 5.7 and Definition 5.8) there also exists $\mathbf{w} \in V$ so that $g(\mathbf{w}, \mathbf{w}) \neq 0$. Applying (3) to $\mathbf{x} = \mathbf{w}$, $\mathbf{z} = \mathbf{u}$, dividing by $g(\mathbf{w}, \mathbf{w})$ proves:

$$(g(\mathbf{u}, \mathbf{w}) - g(\mathbf{w}, \mathbf{u}))g(\mathbf{w}, \mathbf{w}) = 0 \implies g(\mathbf{u}, \mathbf{w}) = g(\mathbf{w}, \mathbf{u}).$$

Similarly applying (3) to $\mathbf{x} = \mathbf{w}$, $\mathbf{z} = \mathbf{v}$, we get $g(\mathbf{v}, \mathbf{w}) = g(\mathbf{w}, \mathbf{v})$. Applying (3) to $\mathbf{x} = \mathbf{v}$, $\mathbf{z} = \mathbf{u}$, dividing by the non-zero quantity $g(\mathbf{u}, \mathbf{v}) - g(\mathbf{v}, \mathbf{u})$ proves $g(\mathbf{v}, \mathbf{v}) = 0$.

Applying (2) to $\mathbf{x} = \mathbf{u}$, $\mathbf{y} = \mathbf{w}$, $\mathbf{z} = \mathbf{v}$,

$$\begin{aligned} g(\mathbf{u}, \mathbf{w})g(\mathbf{v}, \mathbf{u}) - g(\mathbf{u}, \mathbf{v})g(\mathbf{w}, \mathbf{u}) &= 0 \\ \Longrightarrow\ (g(\mathbf{v}, \mathbf{u}) - g(\mathbf{u}, \mathbf{v}))g(\mathbf{u}, \mathbf{w}) &= 0 \implies g(\mathbf{u}, \mathbf{w}) = 0. \end{aligned}$$

Similarly applying (2) to $\mathbf{x} = \mathbf{v}$, $\mathbf{y} = \mathbf{w}$, $\mathbf{z} = \mathbf{u}$ proves $g(\mathbf{v}, \mathbf{w}) = 0$. Using the bi-additive property together with $g(\mathbf{u}, \mathbf{w}) = g(\mathbf{w}, \mathbf{u}) = 0$,

$$\begin{aligned} g(\mathbf{u}, \mathbf{w} + \mathbf{v}) &= g(\mathbf{u}, \mathbf{w}) + g(\mathbf{u}, \mathbf{v}) = g(\mathbf{u}, \mathbf{v}) \\ g(\mathbf{w} + \mathbf{v}, \mathbf{u}) &= g(\mathbf{w}, \mathbf{u}) + g(\mathbf{v}, \mathbf{u}) = g(\mathbf{v}, \mathbf{u}) \\ \Longrightarrow\ g(\mathbf{u}, \mathbf{w} + \mathbf{v}) &\neq g(\mathbf{w} + \mathbf{v}, \mathbf{u}), \end{aligned}$$

and then applying (3) to $\mathbf{x} = \mathbf{w} + \mathbf{v}$, $\mathbf{z} = \mathbf{u}$,

$$(g(\mathbf{u}, \mathbf{w} + \mathbf{v}) - g(\mathbf{w} + \mathbf{v}, \mathbf{u}))g(\mathbf{w} + \mathbf{v}, \mathbf{w} + \mathbf{v}) = 0 \implies g(\mathbf{w} + \mathbf{v}, \mathbf{w} + \mathbf{v}) = 0.$$

However, expanding and simplifying gives:

$$0 = g(\mathbf{w} + \mathbf{v}, \mathbf{w} + \mathbf{v}) = g(\mathbf{w}, \mathbf{w}) + g(\mathbf{w}, \mathbf{v}) + g(\mathbf{v}, \mathbf{w}) + g(\mathbf{v}, \mathbf{v}) = g(\mathbf{w}, \mathbf{w}) + 0 + 0 + 0,$$

which contradicts $g(\mathbf{w}, \mathbf{w}) \neq 0$. The conclusion is that g must be either symmetric or alternating. ∎

Theorem 6.4. *Given a complex vector space V and a function $g : V \times V \to \mathbb{C}$, the following are equivalent:*

1. *g is a Hermitian inner product;*

2. *g is a sesquilinear form and $g(\mathbf{x}, \mathbf{x})$ is real for all $\mathbf{x} \in V$.*

Proof. The proof of *1.* $\implies$ *2.* is left as an easy exercise.

For *2.* $\implies$ *1.*, sesquilinear forms satisfy the first two parts of Definition 6.1, so we only need to check the Hermitian symmetry property. Consider any two vectors $\mathbf{x}$, $\mathbf{y} \in V$. Using the bi-additive property,

$$g(\mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y}) \quad = \quad g(\mathbf{x}, \mathbf{x}) + g(\mathbf{x}, \mathbf{y}) + g(\mathbf{y}, \mathbf{x}) + g(\mathbf{y}, \mathbf{y}),$$

and since $g(\mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y})$ and $g(\mathbf{x}, \mathbf{x}) + g(\mathbf{y}, \mathbf{y})$ are real, $g(\mathbf{x}, \mathbf{y}) + g(\mathbf{y}, \mathbf{x})$ is also real. This real number is equal to its complex conjugate:

$$\begin{aligned} g(\mathbf{x}, \mathbf{y}) + g(\mathbf{y}, \mathbf{x}) &= \overline{g(\mathbf{x}, \mathbf{y}) + g(\mathbf{y}, \mathbf{x})} \\ &= \overline{g(\mathbf{x}, \mathbf{y})} + \overline{g(\mathbf{y}, \mathbf{x})}. \end{aligned} \tag{4}$$

Since (4) holds for all vectors, we can replace $\mathbf{y}$ by $i \cdot \mathbf{y}$ to get another identity, and use the sesquilinear property:

$$\begin{aligned} g(\mathbf{x}, i\cdot\mathbf{y}) + g(i\cdot\mathbf{y}, \mathbf{x}) &= \overline{g(\mathbf{x}, i\cdot\mathbf{y})} + \overline{g(i\cdot\mathbf{y}, \mathbf{x})} \\ \implies (-i)g(\mathbf{x}, \mathbf{y}) + ig(\mathbf{y}, \mathbf{x}) &= \overline{(-i)g(\mathbf{x}, \mathbf{y})} + \overline{ig(\mathbf{y}, \mathbf{x})} \\ &= i\overline{g(\mathbf{x}, \mathbf{y})} - i\overline{g(\mathbf{y}, \mathbf{x})}, \end{aligned}$$

and dividing by $-i$,

$$g(\mathbf{x}, \mathbf{y}) - g(\mathbf{y}, \mathbf{x}) = -\overline{g(\mathbf{x}, \mathbf{y})} + \overline{g(\mathbf{y}, \mathbf{x})}.$$

Adding the last line to Equation (4), there is a cancellation and $2g(\mathbf{x}, \mathbf{y}) = 2\overline{g(\mathbf{y}, \mathbf{x})}$. Dividing by 2 proves the Hermitian symmetry property. ∎

Theorem 6.5. *Given a complex vector space V and a sesquilinear form $g : V \times V \to \mathbb{C}$, the following are equivalent:*

1. *For all $\mathbf{x}, \mathbf{y} \in V$, if $g(\mathbf{x}, \mathbf{y}) = 0$, then $g(\mathbf{y}, \mathbf{x}) = 0$;*

2. *There is some non-zero complex number w so that the function*

$$h(\mathbf{x}, \mathbf{y}) = w \cdot g(\mathbf{x}, \mathbf{y})$$

is a Hermitian inner product on V.

Proof. The proof of *2.* $\implies$ *1.* is left as an easy exercise.

For *1.* $\implies$ *2.*, first consider the case where $g(\mathbf{x}, \mathbf{x}) = 0$ for all $\mathbf{x} \in V$. Then, by Theorem 5.12, g is the constant function 0, so we can let $w = 1$ and $h = g = 0$, and h trivially satisfies the definition of Hermitian inner product.

The only remaining case is that there is some $\mathbf{u} \in V$ so that $g(\mathbf{u}, \mathbf{u}) \neq 0$. Then, let $w = \frac{1}{g(\mathbf{u}, \mathbf{u})}$, so the function $h(\mathbf{x}, \mathbf{y}) = w \cdot g(\mathbf{x}, \mathbf{y})$ satisfies $h(\mathbf{u}, \mathbf{u}) = 1$. Since g is sesquilinear, it is easy to check that h is also sesquilinear. Further, if g satisfies property *1.*, then so does h, using $w \neq 0$, if $h(\mathbf{x},\mathbf{y}) = 0$, then $w \cdot g(\mathbf{x},\mathbf{y}) = 0 \implies g(\mathbf{x},\mathbf{y}) = 0 \implies g(\mathbf{y},\mathbf{x}) = 0 \implies h(\mathbf{y},\mathbf{x}) = w \cdot 0 = 0$.

The following argument will use the assumption that h satisfies *1.* to show that $h(\mathbf{v},\mathbf{v})$ is real for all $\mathbf{v} \in V$. By Theorem 6.4, that would be enough to show that h is a Hermitian inner product, establishing *2.*

The sesquilinear form h satisfies the following identity for any $\mathbf{x},\mathbf{y},\mathbf{z} \in V$:

$$\begin{aligned} h(\mathbf{x}, \overline{h(\mathbf{x},\mathbf{y})} \cdot \mathbf{z} - \overline{h(\mathbf{x},\mathbf{z})} \cdot \mathbf{y}) &= \overline{\overline{h(\mathbf{x},\mathbf{y})}} h(\mathbf{x},\mathbf{z}) - \overline{\overline{h(\mathbf{x},\mathbf{z})}} h(\mathbf{x},\mathbf{y}) \\ &= h(\mathbf{x},\mathbf{y})h(\mathbf{x},\mathbf{z}) - h(\mathbf{x},\mathbf{z})h(\mathbf{x},\mathbf{y}) = 0, \end{aligned}$$

and since h satisfies *1.*, these identities hold:

$$\begin{aligned} h(\overline{h(\mathbf{x},\mathbf{y})} \cdot \mathbf{z} - \overline{h(\mathbf{x},\mathbf{z})} \cdot \mathbf{y}, \mathbf{x}) &= 0 \\ \implies \overline{h(\mathbf{x},\mathbf{y})} h(\mathbf{z},\mathbf{x}) - \overline{h(\mathbf{x},\mathbf{z})} h(\mathbf{y},\mathbf{x}) &= 0. \end{aligned} \tag{5}$$

In the case $\mathbf{y} = \mathbf{x} = \mathbf{u}$, we can use $h(\mathbf{u},\mathbf{u}) = 1$ to get this identity for all $\mathbf{z} \in V$:

$$h(\mathbf{z},\mathbf{u}) - \overline{h(\mathbf{u},\mathbf{z})} = 0. \tag{6}$$

Evaluating Equation (5) with $\mathbf{y} = \mathbf{x}$ and $\mathbf{z} = \mathbf{u}$ gives

$$\overline{h(\mathbf{x},\mathbf{x})} h(\mathbf{u},\mathbf{x}) - \overline{h(\mathbf{x},\mathbf{u})} h(\mathbf{x},\mathbf{x}) = 0,$$

and since $\overline{h(\mathbf{x},\mathbf{u})} = h(\mathbf{u},\mathbf{x})$ by (6), factoring gives

$$(\overline{h(\mathbf{x},\mathbf{x})} - h(\mathbf{x},\mathbf{x}))h(\mathbf{u},\mathbf{x}) = 0. \tag{7}$$

If $\mathbf{x} \in V$ is any vector such that $h(\mathbf{x},\mathbf{x})$ is a non-real complex number, then $\overline{h(\mathbf{x},\mathbf{x})} - h(\mathbf{x},\mathbf{x}) \neq 0$, so by (7), $h(\mathbf{u},\mathbf{x}) = 0$, and also by (6), $h(\mathbf{x},\mathbf{u}) = \overline{h(\mathbf{u},\mathbf{x})} = \overline{0} = 0$.

Suppose, toward a contradiction, that there is some $\mathbf{v} \in V$ so that $h(\mathbf{v},\mathbf{v})$ is a non-real complex number. Then $h(\mathbf{u},\mathbf{v}) = h(\mathbf{v},\mathbf{u}) = 0$, and expanding using the bi-additive property,

$$h(\mathbf{u}+\mathbf{v},\mathbf{u}+\mathbf{v}) = h(\mathbf{u},\mathbf{u}) + h(\mathbf{u},\mathbf{v}) + h(\mathbf{v},\mathbf{u}) + h(\mathbf{v},\mathbf{v}) = 1 + 0 + 0 + h(\mathbf{v},\mathbf{v}),$$

so $\mathbf{u}+\mathbf{v}$ is another vector such that $h(\mathbf{u}+\mathbf{v},\mathbf{u}+\mathbf{v})$ is a non-real complex number, and we can conclude $h(\mathbf{u},\mathbf{u}+\mathbf{v}) = 0$. However, expanding this gives

$$0 = h(\mathbf{u},\mathbf{u}+\mathbf{v}) = h(\mathbf{u},\mathbf{u}) + h(\mathbf{u},\mathbf{v}) = 1 + 0,$$

a contradiction. The conclusion is that there is no element $\mathbf{v} \in V$ so that $h(\mathbf{v},\mathbf{v})$ is a non-real complex number. ∎

The same property *1.* appears in Theorems 6.3 and 6.5 — various authors refer to a bilinear or sesquilinear form satisfying that property as *orthosymmetric* or *reflexive*.

Theorem 6.6. *Given a complex vector space V and a complex symmetric inner product g, if $g(\mathbf{x}, \mathbf{x})$ is real for all $\mathbf{x} \in V$, then g is the constant function zero.*

Proof. Since g is symmetric, it is bilinear, and the property that $g(\mathbf{x}, \mathbf{x})$ is real for all $\mathbf{x}$ implies $g(\mathbf{x}, \mathbf{y}) = -g(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x}$, $\mathbf{y} \in V$ by *4.* $\implies$ *2.* of Theorem 5.10. Using the symmetric property again, $g(\mathbf{x}, \mathbf{y}) = g(\mathbf{y}, \mathbf{x})$, it follows that for all $\mathbf{x}$, $\mathbf{y} \in V$, $g(\mathbf{x}, \mathbf{y}) = -g(\mathbf{x}, \mathbf{y})$, so we can conclude $g(\mathbf{x}, \mathbf{y}) = 0$. ∎

Note that one case in which Theorem 6.6 would apply is where g is positive semidefinite and complex symmetric. The Theorem also shows that the constant function 0 is the only inner product on a complex vector space which is both complex symmetric and Hermitian.

Corollary 6.7. *If V is a complex vector space and $g : V \times V \to \mathbb{C}$ is both complex symmetric and positive definite, then $V = \{\mathbf{0}\}$, the zero-dimensional vector space.*

Proof. By the previous Theorem, g is the constant function 0 since it is symmetric and positive semidefinite. However, since g is positive definite, $g(\mathbf{x}, \mathbf{x}) > 0$ for all non-zero vectors $\mathbf{x} \in V$. The conclusion is that V has no non-zero vectors. ∎

Theorem 6.8. *If g is a symmetric inner product on $\mathbb{K}^n$, then there exists a matrix $G_{n \times n}$ with entries in $\mathbb{K}$ such that*

$$g(\mathbf{x}, \mathbf{y}) = \mathbf{y}^T_{1 \times n} G_{n \times n} \mathbf{x}_{n \times 1},$$

and the matrix $G_{n \times n}$ satisfies $G = G^T$.

Proof. Since g is a bilinear form, the matrix G was constructed in the Proof of Theorem 5.6. From the formula $G_{j\ell} = g(\mathbf{e}^\ell, \mathbf{e}^j)$, and the assumption that g is symmetric, we can conclude $G_{j,\ell} = G_{\ell,j}$, which implies $G = G^T$. ∎

So, we define a "symmetric matrix" as a matrix G equal to its transpose, as in the above Theorem.

Theorem 6.9. *If g is a Hermitian inner product on $\mathbb{K}^n$, then there exists a matrix $G_{n \times n}$ with entries in $\mathbb{K}$ such that*

$$g(\mathbf{x}, \mathbf{y}) = \overline{\mathbf{y}}^T_{1 \times n} G_{n \times n} \mathbf{x},$$

and the matrix $G_{n \times n}$ satisfies $G = \overline{G}^T$.

Proof. Since g is a sesquilinear form, the matrix G was constructed in the Proof of Theorem 5.13. From the formula $G_{j\ell} = g(\mathbf{e}^\ell, \mathbf{e}^j)$, and the assumption that g is Hermitian, we can conclude $G_{j,\ell} = \overline{G_{\ell,j}}$, which implies $G = \overline{G}^T$. ∎

So, we define a "Hermitian matrix" as a matrix G equal to its conjugate transpose, as in the above Theorem. It can have complex entries, but by Proposition 6.2, its diagonal entries, $G_{jj} = g(\mathbf{e}^j, \mathbf{e}^j)$ are real numbers.

Example 6.10. Let V be a complex vector space and let $g : V \times V \to \mathbb{C}$ be an inner product. Then the composite $\mathrm{Re} \circ g : V \times V \to \mathbb{R}$ is a real inner product on V, considered as a real vector space. $\mathrm{Re} \circ g$ is bi-additive, and for real scalars $r \in \mathbb{R}$, $\mathrm{Re}(g(r \cdot \mathbf{x}, \mathbf{y})) = \mathrm{Re}(r \cdot g(\mathbf{x}, \mathbf{y})) = r \cdot \mathrm{Re}(g(\mathbf{x}, \mathbf{y}))$ (the last equality might be false for non-real scalars). In either the case where g is symmetric, or where g is Hermitian, the composite $\mathrm{Re} \circ g$ is symmetric: for all $\mathbf{x}$, $\mathbf{y} \in V$, $\mathrm{Re}(g(\mathbf{x}, \mathbf{y})) = \mathrm{Re}(\overline{g(\mathbf{y}, \mathbf{x})}) = \mathrm{Re}(g(\mathbf{y}, \mathbf{x}))$.

Theorem 6.11 (CBS $\neq$)**.** *Given a positive semidefinite inner product g on a vector space V, for any $\mathbf{x}$, $\mathbf{y} \in V$,*

$$|g(\mathbf{x}, \mathbf{y})|^2 \leq g(\mathbf{x}, \mathbf{x}) g(\mathbf{y}, \mathbf{y}).$$

Proof. Note that if the positive semidefinite inner product g is complex symmetric, then it is the zero function by Theorem 6.6, and the claimed inequality follows trivially. The remaining case is that g is Hermitian (either complex or real).

For any λ, $\mu \in \mathbb{K}$,

$$\begin{aligned} 0 &\leq g(\lambda \cdot \mathbf{x} + \mu \cdot \mathbf{y}, \lambda \cdot \mathbf{x} + \mu \cdot \mathbf{y}) \\ &= \lambda\bar{\lambda} g(\mathbf{x}, \mathbf{x}) + \mu\bar{\lambda} g(\mathbf{y}, \mathbf{x}) + \lambda\bar{\mu} g(\mathbf{x}, \mathbf{y}) + \mu\bar{\mu} g(\mathbf{y}, \mathbf{y}). \end{aligned}$$

In particular, for $\lambda = g(\mathbf{y}, \mathbf{y})$ and $\mu = -g(\mathbf{x}, \mathbf{y})$,

$$\begin{aligned} 0 &\leq \lambda\bar{\lambda} g(\mathbf{x}, \mathbf{x}) + \mu\bar{\lambda}(-\bar{\mu}) + \lambda\bar{\mu}(-\mu) + \mu\bar{\mu}\lambda \\ &= \bar{\lambda}(g(\mathbf{x}, \mathbf{x}) g(\mathbf{y}, \mathbf{y}) - |g(\mathbf{x}, \mathbf{y})|^2), \end{aligned}$$

and if $g(\mathbf{y}, \mathbf{y}) \neq 0$, then this proves the claim. Similarly, for $\lambda = -g(\mathbf{y}, \mathbf{x})$ and $\mu = g(\mathbf{x}, \mathbf{x})$,

$$\begin{aligned} 0 &\leq \lambda\bar{\lambda}\mu + \mu\bar{\lambda}(-\lambda) + \lambda\bar{\mu}(-\bar{\lambda}) + \mu\bar{\mu} g(\mathbf{y}, \mathbf{y}) \\ &= \bar{\mu}(g(\mathbf{x}, \mathbf{x}) g(\mathbf{y}, \mathbf{y}) - |g(\mathbf{y}, \mathbf{x})|^2), \end{aligned}$$

and if $g(\mathbf{x}, \mathbf{x}) \neq 0$, then this proves the claim. Finally, if $g(\mathbf{x}, \mathbf{x}) = g(\mathbf{y}, \mathbf{y}) = 0$, then let $\lambda = 1$ and $\mu = -g(\mathbf{x}, \mathbf{y})$, so

$$\begin{aligned} 0 &\leq 0 - g(\mathbf{x}, \mathbf{y}) g(\mathbf{y}, \mathbf{x}) - g(\mathbf{y}, \mathbf{x}) g(\mathbf{x}, \mathbf{y}) + 0 \\ &= -2|g(\mathbf{x}, \mathbf{y})|^2, \end{aligned}$$

proving $g(\mathbf{x}, \mathbf{y}) = 0$, and the claim. ∎

The following result is a converse of Theorem 5.19 in the case of an inner product. (Note that the bilinear form in Example 5.20 was neither symmetric nor Hermitian.)

Theorem 6.12. *If g is a non-degenerate, positive semidefinite inner product, then g is positive definite.*

Proof. The positive semidefinite property means that the CBS inequality applies, and that $g(\mathbf{x}, \mathbf{x}) \geq 0$ for all $\mathbf{x}$. Suppose that g is not positive definite; then $g(\mathbf{x}, \mathbf{x}) = 0$ for some non-zero $\mathbf{x}$. Then, for any $\mathbf{y}$, $|g(\mathbf{x}, \mathbf{y})|^2 \leq g(\mathbf{x}, \mathbf{x}) g(\mathbf{y}, \mathbf{y}) = 0$, which implies $g(\mathbf{x}, \mathbf{y}) = 0$. This contradicts the assumption that g is non-degenerate. ∎

Theorem 6.13 ($\Delta \neq$)**.** *Given a positive semidefinite inner product g on a vector space V, the function*

$$n_g : V \to \mathbb{R} : n_g(\mathbf{x}) = +\sqrt{g(\mathbf{x}, \mathbf{x})}$$

satisfies, for all $\mathbf{x}$, $\mathbf{y} \in V$,

$$n_g(\mathbf{x} + \mathbf{y}) \leq n_g(\mathbf{x}) + n_g(\mathbf{y}).$$

Proof. It is convenient to denote $n_g(\mathbf{x}) = \|\mathbf{x}\|_g$, keeping track of the inner product g which we're using to define n. We also note that the domain of n is all of V: the real square root is always defined, by the positive semidefinite hypothesis, $g(\mathbf{x},\mathbf{x}) \geq 0$. So, we want to show: $\|\mathbf{x}+\mathbf{y}\|_g \leq \|\mathbf{x}\|_g + \|\mathbf{y}\|_g$.

$$\begin{aligned}
\|\mathbf{x}+\mathbf{y}\|_g^2 &= g(\mathbf{x}+\mathbf{y},\mathbf{x}+\mathbf{y}) \\
&= |g(\mathbf{x},\mathbf{x}) + g(\mathbf{y},\mathbf{x}) + g(\mathbf{x},\mathbf{y}) + g(\mathbf{y},\mathbf{y})| \\
&\leq g(\mathbf{x},\mathbf{x}) + g(\mathbf{y},\mathbf{y}) + 2|g(\mathbf{x},\mathbf{y})| \\
&\leq g(\mathbf{x},\mathbf{x}) + g(\mathbf{y},\mathbf{y}) + 2\sqrt{g(\mathbf{x},\mathbf{x})g(\mathbf{y},\mathbf{y})} \\
&= (\|\mathbf{x}\|_g + \|\mathbf{y}\|_g)^2,
\end{aligned}$$

where the first inequality is the usual triangle inequality for scalar numbers (real or complex), and the second is using the CBS inequality. ∎

Theorem 6.14. *If g is a positive definite inner product on the vector space V, then the function $n_g : V \to \mathbb{R} : n_g(\mathbf{x}) = +\sqrt{g(\mathbf{x},\mathbf{x})} = \|\mathbf{x}\|_g$ is a norm on V.*

Proof. In the case where g is a complex symmetric inner product, $V = \{\mathbf{0}\}$ by Corollary 6.7, so n_g is the constant function 0, which does count as a norm in this case, satisfying Definition 4.6.

In the remaining case where g is (real or complex) Hermitian, it is also easy to check the properties in Definition 4.6 to show n_g is a norm. For $\lambda \in \mathbb{K}$,

$$\|\lambda \cdot \mathbf{x}\|_g = \sqrt{g(\lambda \cdot \mathbf{x}, \lambda \cdot \mathbf{x})} = \sqrt{\lambda\bar{\lambda}g(\mathbf{x},\mathbf{x})} = \sqrt{\lambda\bar{\lambda}} \cdot \|\mathbf{x}\|_g = |\lambda| \cdot \|\mathbf{x}\|_g.$$

The norm function is non-negative by definition, and equals zero only if $\mathbf{x} = \mathbf{0}$, by the fact that g is positive definite. The positive definite property also means that Theorem 6.13 applies, so that the norm satisfies its version of the triangle inequality. ∎

Corollary 6.15. *If g is a positive definite inner product on the vector space V, then the formula*

$$d_g(\mathbf{x},\mathbf{y}) = n_g(\mathbf{y}-\mathbf{x}) = \|\mathbf{y}-\mathbf{x}\|_g = \sqrt{g(\mathbf{y}-\mathbf{x},\mathbf{y}-\mathbf{x})}$$

defines a distance function d_g on V.

Proof. This follows from Proposition 4.7. ∎

Theorem 6.16. *For a distance function d_g as in the previous Corollary, the following identity holds for any $\mathbf{x},\mathbf{y},\mathbf{t} \in V$:*

$$d_g(\mathbf{x}+\mathbf{t},\mathbf{y}+\mathbf{t}) = d_g(\mathbf{x},\mathbf{y}).$$

Proof.

$$\begin{aligned}
d_g(\mathbf{x}+\mathbf{t},\mathbf{y}+\mathbf{t}) &= \sqrt{g((\mathbf{y}+\mathbf{t})-(\mathbf{x}+\mathbf{t}),(\mathbf{y}+\mathbf{t})-(\mathbf{x}+\mathbf{t}))} \\
&= \sqrt{g(\mathbf{y}-\mathbf{x},\mathbf{y}-\mathbf{x})} \\
&= d_g(\mathbf{x},\mathbf{y}).
\end{aligned}$$

∎

This shows that a "translation" function, $\alpha(\mathbf{x}) = \mathbf{x}+\mathbf{t}$, is a motion of V (the inverse function is $\alpha^{-1}(\mathbf{x}) = \mathbf{x}-\mathbf{t}$).

7 Orthogonal and unitary transformations for non-degenerate inner products

Recall, from Theorems 5.22 and 5.25, the equation

$$g(H(\mathbf{x}), H(\mathbf{y})) = g(\mathbf{x}, \mathbf{y}), \tag{8}$$

for a function $H : V \to V$, and a bilinear or sesquilinear form $g : V \times V \to \mathbb{K}$. Also recall the functions $q(\mathbf{x}) = g(\mathbf{x}, \mathbf{x})$ and $\psi(\mathbf{x}, \mathbf{y}) = g(\mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{x})$ from Definition 5.23 and Notation 5.24.

Theorem 7.1. *Given a finite-dimensional (real or complex) vector space V, and a symmetric, non-degenerate inner product $g : V \times V \to \mathbb{K}$, let q and ψ be as defined previously in terms of g. Then, for any function $H : V \to V$, the following are equivalent:*

1. *H is onto, and for all $\mathbf{x}, \mathbf{y} \in V$, $g(H(\mathbf{x}), H(\mathbf{y})) = g(\mathbf{x}, \mathbf{y})$;*
2. *H is onto, and $H(\mathbf{0}) = \mathbf{0}$, and for all $\mathbf{x}, \mathbf{y} \in V$, $\psi(H(\mathbf{x}), H(\mathbf{y})) = \psi(\mathbf{x}, \mathbf{y})$;*
3. *H is linear, and for all $\mathbf{x} \in V$, $q(H(\mathbf{x})) = q(\mathbf{x})$.*

Proof. The proof of equivalence is shown in three steps.

1. $\implies$ *3.*: Since g is a non-degenerate bilinear form and H is onto, Theorem 5.22 applies, to show H is linear. The identity $q(H(\mathbf{x})) = q(\mathbf{x})$ was also already proved, in Theorem 5.25. (This step did not require the symmetric property of g.)

3. $\implies$ *2.*: Since H is linear, $H(\mathbf{0}) = \mathbf{0}$ follows immediately, and using the linearity and the identity for q:

$$\psi(H(\mathbf{x}), H(\mathbf{y})) = q(H(\mathbf{y}) - H(\mathbf{x})) = q(H(\mathbf{y} - \mathbf{x})) = q(\mathbf{y} - \mathbf{x}) = \psi(\mathbf{x}, \mathbf{y}). \tag{9}$$

To show that the linear function $H : V \to V$ is onto, it is enough to show that the kernel of H is only $\{\mathbf{0}\}$ (this is where the finite-dimensional assumption is used). Suppose that $\mathbf{x}$ is in the kernel, so $H(\mathbf{x}) = \mathbf{0}$. Then, for any $\mathbf{y} \in V$, using (9) and the symmetric property of g,

$$\begin{aligned} q(H(\mathbf{y})) &= \psi(\mathbf{0}, H(\mathbf{y})) = \psi(H(\mathbf{x}), H(\mathbf{y})) = \psi(\mathbf{x}, \mathbf{y}) \\ &= q(\mathbf{y}) - 2g(\mathbf{x}, \mathbf{y}) + q(\mathbf{x}) = q(H(\mathbf{y})) - 2g(\mathbf{x}, \mathbf{y}) + q(H(\mathbf{x})), \end{aligned}$$

and since $q(H(\mathbf{x})) = q(\mathbf{0}) = 0$, it follows that $g(\mathbf{x}, \mathbf{y}) = 0$ for all $\mathbf{y}$. Since g is non-degenerate, $\mathbf{x}$ must be $\mathbf{0}$, which is what we wanted to show.

2. $\implies$ *1.*: Using the symmetric property of g,

$$\begin{aligned} \psi(\mathbf{x}, \mathbf{y}) &= g(\mathbf{y}, \mathbf{y}) - 2g(\mathbf{x}, \mathbf{y}) + g(\mathbf{x}, \mathbf{x}) \\ &= \psi(\mathbf{0}, \mathbf{y}) - 2g(\mathbf{x}, \mathbf{y}) + \psi(\mathbf{0}, \mathbf{x}). \end{aligned}$$

Assuming the second and third parts of hypothesis *2.*, the above quantity is equal to:

$$\begin{aligned} \psi(H(\mathbf{x}), H(\mathbf{y})) &= \psi(\mathbf{0}, H(\mathbf{y})) - 2g(H(\mathbf{x}), H(\mathbf{y})) + \psi(\mathbf{0}, H(\mathbf{x})) \\ &= \psi(H(\mathbf{0}), H(\mathbf{y})) - 2g(H(\mathbf{x}), H(\mathbf{y})) + \psi(H(\mathbf{0}), H(\mathbf{x})) \\ &= \psi(\mathbf{0}, \mathbf{y}) - 2g(H(\mathbf{x}), H(\mathbf{y})) + \psi(\mathbf{0}, \mathbf{x}). \end{aligned}$$

By cancelling the equal terms, we can conclude $g(\mathbf{x}, \mathbf{y}) = g(H(\mathbf{x}), H(\mathbf{y}))$. (This step did not require the non-degeneracy property of g or the onto property of H.) ∎

Definition 7.2. Given a finite-dimensional (real or complex) vector space V with a non-degenerate, symmetric inner product g, $H : V \to V$ is an "orthogonal transformation with respect to g" means: H is a function satisfying any of the three equivalent properties from Theorem 7.1.

The Proof of Theorem 7.1 showed that H must be an invertible linear transformation. In the case $V = \mathbb{K}^n$, we also know that any linear function $\mathbb{K}^n \to \mathbb{K}^n$ has a matrix representation. So, for any orthogonal transformation H, there is some non-singular $n \times n$ matrix A so that $H(\mathbf{x}) = A_{n\times n} \cdot \mathbf{x}$.

Theorem 7.3. *Suppose g is a non-degenerate, symmetric inner product on $V = \mathbb{K}^n$, and let G be the matrix from Theorem 6.8. Suppose H is a orthogonal transformation, with matrix representation $H(\mathbf{x}) = A_{n\times n} \cdot \mathbf{x}$. Then, $G = A^T GA$.*

Proof. The proof is similar to the Proof of Theorem 7.6 (just delete the bar and consider $\mathbb{K}^n$ instead of $\mathbb{C}^n$). ∎

Theorem 7.4. *Given a finite-dimensional complex vector space V, and a non-degenerate Hermitian inner product $g : V \times V \to \mathbb{C}$, let q and ψ be as defined previously in terms of g. Then, for any function $H : V \to V$, the following are equivalent:*

1. *H is onto, and for all $\mathbf{x}, \mathbf{y} \in V$, $g(H(\mathbf{x}), H(\mathbf{y})) = g(\mathbf{x}, \mathbf{y})$;*
2. *H is onto, and for all $\mathbf{x}, \mathbf{y} \in V$, $H(i \cdot \mathbf{x}) = i \cdot H(\mathbf{x})$ and $\psi(H(\mathbf{x}), H(\mathbf{y})) = \psi(\mathbf{x}, \mathbf{y})$;*
3. *H is linear, and for all $\mathbf{x} \in V$, $q(H(\mathbf{x})) = q(\mathbf{x})$.*

Proof. The proof of equivalence is shown in three steps.

1. $\implies$ *3.*: This step is the same as in Theorem 7.1.

3. $\implies$ *2.*: Since H is linear, $H(i \cdot \mathbf{x}) = i \cdot H(\mathbf{x})$ follows immediately, and the property that H preserves ψ is proved in the same way as in the Proof of Theorem 7.1, using Equation (9). To show that the linear function $H : V \to V$ is onto, it is enough to show that the kernel of H is only $\{\mathbf{0}\}$ (this is where the finite-dimensional assumption is used). Suppose that $\mathbf{x}$ is in the kernel, so $H(\mathbf{x}) = \mathbf{0}$, and also $H(i \cdot \mathbf{x}) = i \cdot H(\mathbf{x}) = i \cdot \mathbf{0} = \mathbf{0}$. Then, for any $\mathbf{y} \in V$:

$$\begin{aligned}
(i+1)q(H(\mathbf{y})) &= (i+1)\psi(\mathbf{0}, H(\mathbf{y})) = i\psi(\mathbf{0}, H(\mathbf{y})) + \psi(\mathbf{0}, H(\mathbf{y})) \\
&= i\psi(H(\mathbf{x}), H(\mathbf{y})) + \psi(H(i \cdot \mathbf{x}), H(\mathbf{y})) \\
&= i\psi(\mathbf{x}, \mathbf{y}) + \psi(i \cdot \mathbf{x}, \mathbf{y}) \\
&= i(q(\mathbf{y}) - g(\mathbf{y}, \mathbf{x}) - g(\mathbf{x}, \mathbf{y}) + q(\mathbf{x})) \\
&\quad + q(\mathbf{y}) - g(\mathbf{y}, i \cdot \mathbf{x}) - g(i \cdot \mathbf{x}, \mathbf{y}) + q(i \cdot \mathbf{x}) \\
&= (i+1)q(\mathbf{y}) - 2ig(\mathbf{x}, \mathbf{y}) + iq(\mathbf{x}) + q(i \cdot \mathbf{x}) \\
&= (i+1)q(H(\mathbf{y})) - 2ig(\mathbf{x}, \mathbf{y}) + iq(H(\mathbf{x})) + q(H(i \cdot \mathbf{x})),
\end{aligned}$$

where the terms $ig(\mathbf{y}, \mathbf{x})$ and $g(\mathbf{y}, i \cdot \mathbf{x})$ cancel by the sesquilinear property. Since $q(H(\mathbf{x})) = q(H(i \cdot \mathbf{x})) = q(\mathbf{0}) = 0$, it follows that $g(\mathbf{x}, \mathbf{y}) = 0$ for all $\mathbf{y}$. Since g is non-degenerate, $\mathbf{x}$ must be $\mathbf{0}$, which is what we wanted to show.

2. $\implies$ *1.*: We first show that $H(i \cdot \mathbf{x}) = i \cdot H(\mathbf{x})$ implies $H(\mathbf{0}) = \mathbf{0}$. Plug in $\mathbf{x} = \mathbf{0}$, to get $H(i \cdot \mathbf{0}) = H(\mathbf{0}) = i \cdot H(\mathbf{0})$. Then, subtract and simplify: $H(\mathbf{0}) - i \cdot H(\mathbf{0}) = \mathbf{0} \implies (1 - i) \cdot (H(\mathbf{0})) = \mathbf{0}$, so $H(\mathbf{0}) = \mathbf{0}$ by Theorem 1.17.

Recalling Equation (1),

$$\begin{aligned} \psi(\mathbf{x}, \mathbf{y}) &= g(\mathbf{y}, \mathbf{y}) - g(\mathbf{x}, \mathbf{y}) - g(\mathbf{y}, \mathbf{x}) + g(\mathbf{x}, \mathbf{x}) \\ &= \psi(\mathbf{0}, \mathbf{y}) - g(\mathbf{x}, \mathbf{y}) - g(\mathbf{y}, \mathbf{x}) + \psi(\mathbf{0}, \mathbf{x}). \end{aligned}$$

Assuming that H preserves ψ and using the fact (which we just proved) that H fixes $\mathbf{0}$, the above quantity is equal to:

$$\begin{aligned} \psi(H(\mathbf{x}), H(\mathbf{y})) &= \psi(\mathbf{0}, H(\mathbf{y})) - g(H(\mathbf{x}), H(\mathbf{y})) - g(H(\mathbf{y}), H(\mathbf{x})) + \psi(\mathbf{0}, H(\mathbf{x})) \\ &= \psi(H(\mathbf{0}), H(\mathbf{y})) - g(H(\mathbf{x}), H(\mathbf{y})) - g(H(\mathbf{y}), H(\mathbf{x})) + \psi(H(\mathbf{0}), H(\mathbf{x})) \\ &= \psi(\mathbf{0}, \mathbf{y}) - g(H(\mathbf{x}), H(\mathbf{y})) - g(H(\mathbf{y}), H(\mathbf{x})) + \psi(\mathbf{0}, \mathbf{x}). \end{aligned}$$

By cancelling the equal terms, we can conclude

$$g(\mathbf{x}, \mathbf{y}) + g(\mathbf{y}, \mathbf{x}) = g(H(\mathbf{x}), H(\mathbf{y})) + g(H(\mathbf{y}), H(\mathbf{x})). \tag{10}$$

Since this identity holds for all $\mathbf{x}, \mathbf{y} \in V$, we can substitute $i \cdot \mathbf{x}$ for $\mathbf{x}$ to get this identity:

$$g(i \cdot \mathbf{x}, \mathbf{y}) + g(\mathbf{y}, i \cdot \mathbf{x}) = g(H(i \cdot \mathbf{x}), H(\mathbf{y})) + g(H(\mathbf{y}), H(i \cdot \mathbf{x})).$$

Then, using the assumption that $H(i \cdot \mathbf{x}) = i \cdot H(\mathbf{x})$ and the sesquilinear property of g,

$$\begin{aligned} ig(\mathbf{x}, \mathbf{y}) + (-i)g(\mathbf{y}, \mathbf{x}) &= g(i \cdot H(\mathbf{x}), H(\mathbf{y})) + g(H(\mathbf{y}), i \cdot H(\mathbf{x})) \\ &= ig(H(\mathbf{x}), H(\mathbf{y})) + (-i)g(H(\mathbf{y}), H(\mathbf{x})), \end{aligned}$$

and dividing both sides by i gives

$$g(\mathbf{x}, \mathbf{y}) - g(\mathbf{y}, \mathbf{x}) = g(H(\mathbf{x}), H(\mathbf{y})) - g(H(\mathbf{y}), H(\mathbf{x})).$$

Adding this identity to Equation (10) gives

$$2g(\mathbf{x}, \mathbf{y}) = 2g(H(\mathbf{x}), H(\mathbf{y})),$$

which implies *1.* (This step did not require the non-degeneracy property of g or the onto property of H.) ∎

Definition 7.5. Given a finite-dimensional complex vector space V with a non-degenerate, Hermitian inner product g, $H : V \to V$ is a "unitary transformation with respect to g" means: H is a function satisfying any of the three equivalent properties from Theorem 7.4.

Theorem 7.6. *Suppose g is a non-degenerate, Hermitian inner product on $V = \mathbb{C}^n$, and let G be the matrix from Theorem 6.9. Suppose H is a unitary transformation, with matrix representation $H(\mathbf{x}) = A_{n \times n} \cdot \mathbf{x}$. Then, $G = \bar{A}^T G A$.*

Proof. G is invertible by Theorem 5.18. Let $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$.

$$\begin{aligned} g(G^{-1}\bar{A}^T GA\mathbf{x}, \mathbf{y}) &= \bar{\mathbf{y}}^T GG^{-1}\bar{A}^T GA\mathbf{x} \\ &= \bar{\mathbf{y}}^T \bar{A}^T GA\mathbf{x} \\ &= \overline{(A\mathbf{y})}^T GA\mathbf{x} \\ &= g(A\mathbf{x}, A\mathbf{y}) = g(H(\mathbf{x}), H(\mathbf{y})) \\ &= g(\mathbf{x}, \mathbf{y}), \end{aligned}$$

so by Theorem 5.21, we can conclude $G^{-1}\bar{A}^T GA\mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$, so $G^{-1}\bar{A}^T GA = I_{n\times n}$ (the identity matrix), and the result follows from multiplying both sides by G. ∎

In analogy with Definitions 7.2 and 7.5, we have the following terms for matrices (with entries in $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$).

Definition 7.7. Given an invertible symmetric matrix $G_{n\times n}$, a matrix $A_{n\times n}$ which satisfies the equation $G = A^T GA$ is called "orthogonal with respect to G."

Definition 7.8. Given an invertible Hermitian matrix $G_{n\times n}$, a matrix A which satisfies the equation $G = \bar{A}^T GA$ is called "unitary with respect to G."

When G and A are real matrices, "orthogonal" and "unitary" mean the same thing. In the above two Definitions, it follows that A is non-singular (for example, by taking the determinant of both sides).

Example 7.9. Let G be the identity matrix $I_{n\times n}$, which is symmetric. The symmetric inner product on $\mathbb{R}^n$ given by the formula

$$g(\mathbf{x}, \mathbf{y}) = \mathbf{y}^T G\mathbf{x} = \mathbf{y}^T\mathbf{x} = x_1y_1 + x_2y_2 + \cdots + x_ny_n$$

is called the "dot product" or the "real Euclidean inner product." It is positive definite. A real matrix A which is orthogonal with respect to $G = I$ satisfies the equation $A^T IA = A^T A = I$, or equivalently, $A^T = A^{-1}$.

Example 7.10. Let G be the identity matrix $I_{n\times n}$, which is symmetric. The symmetric inner product on $\mathbb{C}^n$ given by the formula

$$g(\mathbf{x}, \mathbf{y}) = \mathbf{y}^T G\mathbf{x} = \mathbf{y}^T\mathbf{x} = x_1y_1 + x_2y_2 + \cdots + x_ny_n$$

is called the "complex symmetric dot product." For $n > 0$, it is not positive definite, by Theorem 6.6, but since I is invertible, g is non-degenerate, by Theorem 5.18. A complex matrix A which is orthogonal with respect to $G = I$ satisfies the equation $A^T IA = A^T A = I$, or equivalently, $A^T = A^{-1}$.

Example 7.11. Let G be the identity matrix $I_{n\times n}$, which is Hermitian. The Hermitian inner product on $\mathbb{C}^n$ given by the formula

$$g(\mathbf{x}, \mathbf{y}) = \bar{\mathbf{y}}^T G\mathbf{x} = \bar{\mathbf{y}}^T\mathbf{x} = x_1\bar{y}_1 + x_2\bar{y}_2 + \cdots + x_n\bar{y}_n$$

is called the "Hermitian dot product" or the "complex Euclidean inner product." It is positive definite. A complex matrix A which is unitary with respect to $G = I$ satisfies the equation $\bar{A}^T IA = \bar{A}^T A = I$, or equivalently, $\bar{A}^T = A^{-1}$.

8 Orthogonal and unitary transformations for positive definite inner products

Recall, from Theorem 6.14 and Corollary 6.15, the norm function $n_g : V \to \mathbb{R} : n_g(\mathbf{x}) = \sqrt{g(\mathbf{x}, \mathbf{x})}$ and distance function $d_g(\mathbf{x}, \mathbf{y}) = \sqrt{g(\mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{x})}$ defined in terms of a positive definite inner product g. For a real vector space with a positive definite inner product, we can drop the "onto" assumptions from Theorem 7.1, although the Proof will be different.

Theorem 8.1. *Given a finite-dimensional real vector space V, and a positive definite inner product $g : V \times V \to \mathbb{R}$, let n_g and d_g be the norm and distance functions defined by g as above. Then, for any function $H : V \to V$, the following are equivalent:*

1. *For all $\mathbf{x}, \mathbf{y} \in V$, $g(H(\mathbf{x}), H(\mathbf{y})) = g(\mathbf{x}, \mathbf{y})$;*
2. *$H(\mathbf{0}) = \mathbf{0}$, and for all $\mathbf{x}, \mathbf{y} \in V$, $d_g(H(\mathbf{x}), H(\mathbf{y})) = d_g(\mathbf{x}, \mathbf{y})$;*
3. *H is linear, and for all $\mathbf{x} \in V$, $n_g(H(\mathbf{x})) = n_g(\mathbf{x})$;*
4. *H is an orthogonal transformation of V with respect to g;*
5. *H is a motion of V, and $H(\mathbf{0}) = \mathbf{0}$.*

Proof. Let $\psi(\mathbf{x}, \mathbf{y}) = (d_g(\mathbf{x}, \mathbf{y}))^2 = g(\mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{x})$ and $q(\mathbf{x}) = (n_g(\mathbf{x}))^2 = g(\mathbf{x}, \mathbf{x})$ as in Theorem 7.1, which applies here since an inner product on a real vector space is symmetric, and a positive definite inner product is non-degenerate. Then $d_g(H(\mathbf{x}), H(\mathbf{y})) = d_g(\mathbf{x}, \mathbf{y}) \iff \psi(H(\mathbf{x}), H(\mathbf{y})) = \psi(\mathbf{x}, \mathbf{y})$, by taking the non-negative square root, and similarly, $n_g(H(\mathbf{x})) = n_g(\mathbf{x}) \iff q(H(\mathbf{x})) = q(\mathbf{x})$. So, *3.* of this Theorem implies *1.* and *2.* of this Theorem by Theorem 7.1. Since *3.* also implies H is onto (as in Theorem 7.1), *3.* $\implies$ *5.* $\implies$ *2.* by Definition 4.4.

To show *1.* $\implies$ *2.*, first consider $g(H(\mathbf{0}), H(\mathbf{0})) = g(\mathbf{0}, \mathbf{0}) = 0$; the positive definite property of g implies $H(\mathbf{0}) = \mathbf{0}$. The property $\psi(H(\mathbf{x}), H(\mathbf{y})) = \psi(\mathbf{x}, \mathbf{y})$ was proved in Theorem 5.25, which implies *2.*

Since we have *3.* $\implies$ *1.* $\implies$ *2.* and *3.* $\implies$ *5.* $\implies$ *2.*, and *3.* $\iff$ *4.* by Definition 7.2, the only remaining step is to show *2.* $\implies$ *3.* The fact that *2.* implies $n_g(H(\mathbf{x})) = n_g(\mathbf{x})$ is easy: $n_g(H(\mathbf{x})) = d_g(\mathbf{0}, H(\mathbf{x})) = d_g(H(\mathbf{0}), H(\mathbf{x})) = d_g(\mathbf{0}, \mathbf{x}) = n_g(\mathbf{x})$. Showing that H is linear uses some tricky identities:

From Exercise 5.4:

$$\begin{aligned}
& (d_g(\mathbf{u} + \mathbf{v}, \mathbf{w}))^2 \qquad (11)\\
= \ & g(\mathbf{u} + \mathbf{v} - \mathbf{w}, \mathbf{u} + \mathbf{v} - \mathbf{w}) \\
= \ & g(\mathbf{u} - \mathbf{w}, \mathbf{u} - \mathbf{w}) + g(\mathbf{v} - \mathbf{w}, \mathbf{v} - \mathbf{w}) - g(\mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v}) + g(\mathbf{u}, \mathbf{u}) + g(\mathbf{v}, \mathbf{v}) - g(\mathbf{w}, \mathbf{w}) \\
= \ & (d_g(\mathbf{u}, \mathbf{w}))^2 + (d_g(\mathbf{v}, \mathbf{w}))^2 - (d_g(\mathbf{u}, \mathbf{v}))^2 + (n_g(\mathbf{u}))^2 + (n_g(\mathbf{v}))^2 - (n_g(\mathbf{w}))^2.
\end{aligned}$$

Since H preserves both d_g and n_g, this last quantity is equal to:

$$\begin{aligned}
& (d_g(H(\mathbf{u}), H(\mathbf{w})))^2 + (d_g(H(\mathbf{v}), H(\mathbf{w})))^2 - (d_g(H(\mathbf{u}), H(\mathbf{v})))^2 \\
& +(n_g(H(\mathbf{u})))^2 + (n_g(H(\mathbf{v})))^2 - (n_g(H(\mathbf{w})))^2, \\
= \ & (d_g(H(\mathbf{u}) + H(\mathbf{v}), H(\mathbf{w})))^2,
\end{aligned}$$

the last step using Exercise 5.4 again. Setting $\mathbf{w} = \mathbf{u}+\mathbf{v}$, line (11) is 0, so $d_g(H(\mathbf{u})+H(\mathbf{v}), H(\mathbf{u}+\mathbf{v})) = 0$, and we can conclude $H(\mathbf{u}) + H(\mathbf{v}) = H(\mathbf{u} + \mathbf{v})$.

From Exercise 5.5:

$$\begin{aligned} & (d_g(\lambda \cdot \mathbf{u}, \mathbf{w}))^2 \\ = \;& g(\lambda \cdot \mathbf{u} - \mathbf{w}, \lambda \cdot \mathbf{u} - \mathbf{w}) \\ = \;& (1-\lambda)(g(\mathbf{w}, \mathbf{w}) - \lambda g(\mathbf{u}, \mathbf{u})) + \lambda g(\mathbf{u} - \mathbf{w}, \mathbf{u} - \mathbf{w}) \\ = \;& (1-\lambda)((n_g(\mathbf{w}))^2 - \lambda(n_g(\mathbf{u}))^2) + \lambda(d_g(\mathbf{u}, \mathbf{w}))^2. \end{aligned} \tag{12}$$

Since H preserves both d_g and n_g, this last quantity is equal to:

$$\begin{aligned} & (1-\lambda)((n_g(H(\mathbf{w})))^2 - \lambda(n_g(H(\mathbf{u})))^2) + \lambda(d_g(H(\mathbf{u}), H(\mathbf{w})))^2 \\ = \;& (d_g(\lambda \cdot H(\mathbf{u}), H(\mathbf{w})))^2, \end{aligned}$$

the last step using Exercise 5.5 again. Setting $\mathbf{w} = \lambda \cdot \mathbf{u}$, line (12) is 0, so $d_g(\lambda \cdot H(\mathbf{u}), H(\lambda \cdot \mathbf{u})) = 0$, and we can conclude $H(\lambda \cdot \mathbf{u}) = \lambda \cdot H(\mathbf{u})$. ∎

Corollary 8.2. *Given a finite-dimensional real vector space V, and a positive definite inner product $g : V \times V \to \mathbb{R}$, d_g be the distance function defined by g as above. Then, for any function $M : V \to V$, the following are equivalent:*

1. *M is a motion of V;*
2. *For all $\mathbf{x}, \mathbf{y} \in V$, $d_g(M(\mathbf{x}), M(\mathbf{y})) = d_g(\mathbf{x}, \mathbf{y})$;*
3. *There exists a vector $\mathbf{t} \in V$ and a function $H : V \to V$ which is an orthogonal transformation of V with respect to g, and such that for all $\mathbf{x} \in V$, $M(\mathbf{x}) = H(\mathbf{x}) + \mathbf{t}$.*

Proof. *1.* $\implies$ *2.* by Definition 4.4.

2. $\implies$ *3.*: If M preserves distances, then define $\mathbf{t} = M(\mathbf{0})$ and $H(\mathbf{x}) = M(\mathbf{x}) - M(\mathbf{0})$, so that $H(\mathbf{0}) = M(\mathbf{0}) - M(\mathbf{0}) = \mathbf{0}$. H preserves distances, using Theorem 6.16:

$$\begin{aligned} d_g(H(\mathbf{x}), H(\mathbf{y})) &= d_g(M(\mathbf{x}) - \mathbf{t}, M(\mathbf{y}) - \mathbf{t}) \\ &= d_g(M(\mathbf{x}), M(\mathbf{y})) = d_g(\mathbf{x}, \mathbf{y}). \end{aligned}$$

So, H is an orthogonal transformation by Theorem 8.1.

3. $\implies$ *1.*: The function $M(\mathbf{x}) = H(\mathbf{x}) + \mathbf{t}$ preserves distances, using Theorem 6.16 again:

$$\begin{aligned} d_g(M(\mathbf{x}), M(\mathbf{y})) &= d_g(H(\mathbf{x}) + \mathbf{t}, H(\mathbf{y}) + \mathbf{t}) \\ &= d_g(H(\mathbf{x}), H(\mathbf{y})) = d_g(\mathbf{x}, \mathbf{y}). \end{aligned}$$

Also, M is onto because H is onto (by Theorem 7.1): for any $\mathbf{y} \in V$, there is some $\mathbf{x} \in V$ so that $H(\mathbf{x}) = \mathbf{y} - \mathbf{t}$, so $\mathbf{y} = H(\mathbf{x}) + \mathbf{t} = M(\mathbf{x})$. ∎

Corollary 8.3. *Given a positive definite real inner product g on $\mathbb{R}^n$, let G be the matrix corresponding to g from Theorem 6.8, and let d_g be the distance function corresponding to g as above. Then, for any function $M : \mathbb{R}^n \to \mathbb{R}^n$, the following are equivalent:*

1. *M is a motion of $\mathbb{R}^n$;*
2. *For all $\vec{x}, \vec{y} \in \mathbb{R}^n$, $d_g(M(\vec{x}), M(\vec{y})) = d_g(\vec{x}, \vec{y})$;*
3. *There exists a vector $\vec{t} \in \mathbb{R}^n$ and a $n \times n$ real matrix A so that $G = A^T G A$, and for all $\vec{x} \in \mathbb{R}^n$, $M(\vec{x}) = A_{n\times n} \cdot \vec{x}_{n\times 1} + \vec{t}_{n\times 1}$.*

Proof. Let H be the orthogonal transformation from the previous Corollary. The representation of the function H by a matrix A with the claimed property was established in Theorem 7.3, so A is an orthogonal matrix, as in Definition 7.7. ∎

This result shows that if M is any transformation from $\mathbb{R}^n$ to itself that preserves the distance function d_g, then M has to be equal to matrix multiplication by an orthogonal matrix (for example, a rotation or a reflection), followed by a translation (vector addition of $\vec{t}$). This characterization of distance-preserving functions applies only when the distance is defined in terms of an inner product on a vector space, not necessarily to other types of distance functions on $\mathbb{R}^n$ (or on other sets).

Exercise 8.4. Show that a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ which is orthogonal with respect to a positive definite inner product on $\mathbb{R}^3$ has at least one "eigenvector" $\vec{v}$ with "eigenvalue" either 1 or -1.

Hint. Let $A_{3\times 3}$ be a real matrix representing T: $T(\vec{v}) = A\vec{v}$. The problem is to show there exists a non-zero vector $\vec{v} \in \mathbb{R}^3$ such that either $A\vec{v} = \vec{v}$ or $A\vec{v} = -\vec{v}$. Show first that there exists an eigenvector, and then show that the eigenvalue must be ± 1 using the fact that the inner product is preserved. ∎

Theorem 8.5. *Given a finite-dimensional complex vector space V, and a positive definite inner product $g : V \times V \to \mathbb{C}$, let n_g and d_g be the norm and distance functions defined by g. Then, for any function $H : V \to V$, the following are equivalent:*

1. *For all $\mathbf{x}, \mathbf{y} \in V$, $\mathrm{Re}(g(H(\mathbf{x}), H(\mathbf{y}))) = \mathrm{Re}(g(\mathbf{x}, \mathbf{y}))$;*
2. *$H(\mathbf{0}) = \mathbf{0}$, and for all $\mathbf{x}, \mathbf{y} \in V$, $d_g(H(\mathbf{x}), H(\mathbf{y})) = d_g(\mathbf{x}, \mathbf{y})$;*
3. *H is additive, satisfies $H(r \cdot \mathbf{x}) = r \cdot H(\mathbf{x})$ for all $r \in \mathbb{R}$, and for all $\mathbf{x} \in V$, $n_g(H(\mathbf{x})) = n_g(\mathbf{x})$;*
4. *H is an orthogonal transformation of V (considered as a real vector space) with respect to the real inner product $\mathrm{Re} \circ g$;*
5. *H is a motion of V with respect to d_g, and $H(\mathbf{0}) = \mathbf{0}$.*

Proof. If g is complex symmetric and positive definite, then $V = \{\mathbf{0}\}$ by Corollary 6.7, so H is the constant function zero, which satisfies all of the above properties.

In the case where g is Hermitian and positive definite, the composite function $\mathrm{Re} \circ g$ is a positive definite real inner product on the real vector space V, as in Examples 5.29 and 6.10.

Since V is a finite-dimensional complex vector space (with some basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$), it is also finite-dimensional considered as a real vector space (although in general we need twice as many basis vectors to span V using only real coefficients, for example, it is straightforward to check $\{\mathbf{v}_1, \ldots, \mathbf{v}_n, i \cdot \mathbf{v}_1, \ldots, i \cdot \mathbf{v}_n\}$ is a basis for the real vector space V).

The two inner products g and $\mathrm{Re} \circ g$ define exactly the same distance function, $d_g = d_{\mathrm{Re} \circ g}$:

$$d_{\mathrm{Re} \circ g}(\mathbf{x}, \mathbf{y}) = \sqrt{\mathrm{Re}(g(\mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{x}))} = \sqrt{g(\mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{x})} = d_g(\mathbf{x}, \mathbf{y}).$$

So, a function H preserves the d_g distance as in *2.* if and only if H preserves the $d_{\mathrm{Re} \circ g}$ distance, and this Theorem will follow from applying Theorem 8.1 to the real vector space V with the real symmetric positive definite inner product $\mathrm{Re} \circ g$. Specifically, statement *2.* of this Theorem referring to d_g is equivalent to *2.* from Theorem 8.1 referring to $d_{\mathrm{Re} \circ g}$, which is equivalent to H being a real linear orthogonal transformation of the real vector space V (*4.*), that preserves the inner product $\mathrm{Re} \circ g$ (*1.*) and the norm $n_{\mathrm{Re} \circ g} = n_g$ (*3.*), and which is a motion of V with respect to $d_{\mathrm{Re} \circ g}$, or equivalently with respect to the same distance function d_g (*5.*). ∎

Corollary 8.6. *Given a finite-dimensional complex vector space V, and a positive definite inner product $g : V \times V \to \mathbb{C}$, let d_g be the distance function defined by g. Then, for any function $M : V \to V$, the following are equivalent:*

1. *M is a motion of V;*
2. *For all $\mathbf{x}, \mathbf{y} \in V$, $d_g(M(\mathbf{x}), M(\mathbf{y})) = d_g(\mathbf{x}, \mathbf{y})$;*
3. *There exists a vector $\mathbf{t} \in V$ and a function $H : V \to V$ which is an orthogonal transformation of the real vector space V with respect to the real inner product $\mathrm{Re} \circ g$, and such that for all $\mathbf{x} \in V$, $M(\mathbf{x}) = H(\mathbf{x}) + \mathbf{t}$.*

Proof. This follows from the previous Theorem in the same way that Corollary 8.2 followed from Theorem 8.1. ∎

Example 8.7. Consider the complex vector space $V = \mathbb{C}^2$, with the Hermitian dot product from Example 7.11: for $\mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$, $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$,

$$g(\mathbf{z}, \mathbf{w}) = z_1 \bar{w}_1 + z_2 \bar{w}_2 = \overline{\mathbf{w}}^T \mathbf{z}.$$

The following function is an example of a distance-preserving map from V onto V: let

$$H(\mathbf{z}) = \begin{pmatrix} \bar{z}_1 \\ \bar{z}_2 \end{pmatrix},$$

so H applies complex conjugation to both components of $\mathbf{z}$. It is easy to check that H is additive, and satisfies $H(r \cdot \mathbf{z}) = r \cdot H(\mathbf{z})$ for real scalars r; but it is also easy to show (by example) that H is not linear because $H(i \cdot \mathbf{z}) \neq i \cdot H(\mathbf{z})$, and H does not preserve the complex Hermitian inner product g. In fact, $g(H(\mathbf{z}), H(\mathbf{w})) = \overline{g(\mathbf{z}, \mathbf{w})}$, and these quantities have the same real part, so H preserves the real symmetric inner product $\mathrm{Re} \circ g$. This function H is an example of a non-unitary function satisfying the equivalent conditions of Theorem 8.5, and also explains why the hypothesis $H(i \cdot \mathbf{x}) = i \cdot H(\mathbf{x})$ that was needed in part *2.* of Theorem 7.4 is different from the hypothesis $H(\mathbf{0}) = \mathbf{0}$ from part *2.* of Theorem 7.1.

Theorem 8.8. *Given a finite-dimensional complex vector space V, and a positive definite inner product $g : V \times V \to \mathbb{C}$, let n_g and d_g be the norm and distance functions defined by g. Then, for any function $H : V \to V$, the following are equivalent:*

1. *For all $\mathbf{x}, \mathbf{y} \in V$, $g(H(\mathbf{x}), H(\mathbf{y})) = g(\mathbf{x}, \mathbf{y})$;*
2. *For all $\mathbf{x}, \mathbf{y} \in V$, $H(i \cdot \mathbf{x}) = i \cdot H(\mathbf{x})$ and $d_g(H(\mathbf{x}), H(\mathbf{y})) = d_g(\mathbf{x}, \mathbf{y})$;*
3. *H is linear, and for all $\mathbf{x} \in V$, $n_g(H(\mathbf{x})) = n_g(\mathbf{x})$;*
4. *H is a unitary transformation of V with respect to g;*
5. *H is a motion of V such that for all $\mathbf{x} \in V$, $H(i \cdot \mathbf{x}) = i \cdot H(\mathbf{x})$.*

Proof. As in the Proof of Theorem 8.5, if g is complex symmetric and positive definite, then H is the constant function $\mathbf{0}$ on $V = \{\mathbf{0}\}$, satisfying all the equivalent properties. So we continue by considering the case where g is Hermitian.

Let $\psi(\mathbf{x}, \mathbf{y}) = (d_g(\mathbf{x}, \mathbf{y}))^2 = g(\mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{x})$ and $q(\mathbf{x}) = (n_g(\mathbf{x}))^2 = g(\mathbf{x}, \mathbf{x})$ as in Theorem 7.4, which applies here since a positive definite inner product is non-degenerate. Then $d_g(H(\mathbf{x}), H(\mathbf{y})) = d_g(\mathbf{x}, \mathbf{y}) \iff \psi(H(\mathbf{x}), H(\mathbf{y})) = \psi(\mathbf{x}, \mathbf{y})$, by taking the non-negative square root, and similarly, $n_g(H(\mathbf{x})) = n_g(\mathbf{x}) \iff q(H(\mathbf{x})) = q(\mathbf{x})$. So, *3.* of this Theorem implies *1.* and *2.* of this Theorem by Theorem 7.4. Since *3.* also implies H is onto (as in Theorem 7.4), *3.* $\implies$ *5.* $\implies$ *2.* by Definition 4.4.

To show *1.* $\implies$ *2.*, the distance-preserving property follows in the same way as in the Proof of Theorem 8.1. To show $H(i \cdot \mathbf{x}) = i \cdot H(\mathbf{x})$, we use the sesquilinear and positive definite properties of g. For any $\mathbf{u}$, $\mathbf{w} \in V$,

$$\begin{aligned} & g(i \cdot \mathbf{u} - \mathbf{w}, i \cdot \mathbf{u} - \mathbf{w}) \\ = \ & g(\mathbf{u}, \mathbf{u}) - ig(\mathbf{u}, \mathbf{w}) + ig(\mathbf{w}, \mathbf{u}) + g(\mathbf{w}, \mathbf{w}) \\ = \ & g(H(\mathbf{u}), H(\mathbf{u})) - ig(H(\mathbf{u}), H(\mathbf{w})) + ig(H(\mathbf{w}), H(\mathbf{u})) + g(H(\mathbf{w}), H(\mathbf{w})) \\ = \ & g(i \cdot H(\mathbf{u}) - H(\mathbf{w}), i \cdot H(\mathbf{u}) - H(\mathbf{w})). \end{aligned}$$

So if $\mathbf{w} = i \cdot \mathbf{u}$, then LHS $= 0 =$ RHS, and we can conclude $i \cdot H(\mathbf{u}) - H(i \cdot \mathbf{u}) = \mathbf{0}$.

Since we have *3.* $\implies$ *1.* $\implies$ *2.* and *3.* $\implies$ *5.* $\implies$ *2.*, and *3.* $\iff$ *4.* by Definition 7.5, the only remaining step is to show *2.* $\implies$ 3. Exactly as in the Proof of Theorem 7.4, $H(i \cdot \mathbf{x}) = i \cdot H(\mathbf{x})$ implies $H(\mathbf{0}) = \mathbf{0}$, so statement *2.* of this Theorem implies statement *2.* of Theorem 8.5, which implies statement *3.* of that Theorem. We can conclude $n_g(H(\mathbf{x})) = n_g(\mathbf{x})$, and also that H is additive and satisfies $H(r \cdot \mathbf{x}) = r \cdot H(\mathbf{x})$ for all $r \in \mathbb{R}$. To show H is linear, we use these properties together with the additional assumption that $H(i \cdot \mathbf{x}) = i \cdot H(\mathbf{x})$. For any $z \in \mathbb{C}$, let $z = x + iy$. Then

$$\begin{aligned} H(z \cdot \mathbf{u}) &= H((x + iy) \cdot \mathbf{u}) = H((x \cdot \mathbf{u}) + ((iy) \cdot \mathbf{u})) = H(x \cdot \mathbf{u}) + H((iy) \cdot \mathbf{u}) \\ &= x \cdot H(\mathbf{u}) + H(i \cdot (y \cdot \mathbf{u})) = x \cdot H(\mathbf{u}) + i \cdot H(y \cdot \mathbf{u}) = x \cdot H(\mathbf{u}) + i \cdot (y \cdot H(\mathbf{u})) \\ &= x \cdot H(\mathbf{u}) + (iy) \cdot H(\mathbf{u}) = (x + iy) \cdot H(\mathbf{u}) = z \cdot H(\mathbf{u}). \end{aligned}$$

∎

Corollary 8.9. *Given a finite-dimensional complex vector space V, and a positive definite inner product $g : V \times V \to \mathbb{C}$, let d_g be the distance function defined by g. Then, for any function $M : V \to V$, the following are equivalent:*

1. *M is a motion of V such that for all $\mathbf{x} \in V$, $M(i \cdot \mathbf{x}) - M(\mathbf{0}) = i \cdot (M(\mathbf{x}) - M(\mathbf{0}))$;*
2. *For all $\mathbf{x}, \mathbf{y} \in V$, $M(i \cdot \mathbf{x}) - M(\mathbf{0}) = i \cdot (M(\mathbf{x}) - M(\mathbf{0}))$ and $d_g(M(\mathbf{x}), M(\mathbf{y})) = d_g(\mathbf{x}, \mathbf{y})$;*
3. *There exists a vector $\mathbf{t} \in V$ and a function $H : V \to V$ which is a unitary transformation of the complex vector space V with respect to the inner product g, and such that for all $\mathbf{x} \in V$, $M(\mathbf{x}) = H(\mathbf{x}) + \mathbf{t}$.*

Proof. This follows from the previous Theorem in the same way that Corollary 8.2 followed from Theorem 8.1. The construction of $H(\mathbf{x}) = M(\mathbf{x}) - M(\mathbf{0})$ shows that the condition $M(i \cdot \mathbf{x}) - M(\mathbf{0}) = i \cdot (M(\mathbf{x}) - M(\mathbf{0}))$ is equivalent to $H(i \cdot \mathbf{x}) = i \cdot H(\mathbf{x})$. ∎

Corollary 8.10. *Given a positive definite inner product g on $\mathbb{C}^n$, let G be the complex matrix corresponding to g from Theorem 6.9, and let d_g be the distance function defined by g. Then, for any function $M : \mathbb{C}^n \to \mathbb{C}^n$, the following are equivalent:*

1. *M is a motion of $\mathbb{C}^n$ such that for all $\vec{x} \in \mathbb{C}^n$, $M(i \cdot \vec{x}) - M(\vec{0}) = i \cdot (M(\vec{x}) - M(\vec{0}))$;*
2. *For all $\vec{x}, \vec{y} \in \mathbb{C}^n$, $M(i \cdot \vec{x}) - M(\vec{0}) = i \cdot (M(\vec{x}) - M(\vec{0}))$ and $d_g(M(\vec{x}), M(\vec{y})) = d_g(\vec{x}, \vec{y})$;*
3. *There exists a vector $\vec{t} \in \mathbb{C}^n$ and a $n \times n$ complex matrix A such that $G = \bar{A}^T G A$, and for all $\vec{x} \in \mathbb{C}^n$, $M(\vec{x}) = A \cdot \vec{x} + \vec{t}$.*

Proof. Let H be the unitary transformation from the previous Corollary. The representation of the function H by a matrix A with the claimed property was established in Theorem 7.6, so A is a unitary matrix, as in Definition 7.8. ∎

The following Theorem shows that a mapping of a spanning subset which preserves inner products extends to a unitary transformation.

Theorem 8.11. *Given a finite-dimensional complex vector space V, with a positive definite inner product g, and a subset $S \subseteq V$ such that the span of S is V, suppose there is a function $T : S \to V$ such that $g(\vec{v}, \vec{w}) = g(T(\vec{v}), T(\vec{w}))$ for all $\vec{v}, \vec{w} \in S$. Then there is a function $H : V \to V$ such that $H(\vec{v}) = T(\vec{v})$ for all $\vec{v} \in S$, and H is unitary.*

Proof. By Theorem 8.8, the unitary property will follow from showing $g(\vec{v}, \vec{w}) = g(H(\vec{v}), H(\vec{w}))$ for all $\vec{v}, \vec{w} \in V$.

Define H as follows: for $\vec{v} \in V$, the spanning property of S means that $\vec{v} = \sum c_i \vec{v}_i$, for finitely many $\{\vec{v}_1, \ldots, \vec{v}_n\} \subseteq S$. Then define $H(\vec{v}) = \sum_{i=1}^{n} c_i T(\vec{v}_i)$. However, since we are not assuming S is an independent set (it may in fact be infinite), $\vec{v}$ may also be expressible as some other linear combination: $\vec{v} = \sum_{i=1}^{N} d_i \vec{v}_i$, for a possibly longer, but still finite, list $\{\vec{v}_1, \ldots, \vec{v}_n, \ldots, \vec{v}_N\}$. To show H is well-defined, we need to show $\sum_{i=1}^{N} c_i T(\vec{v}_i) = \sum_{i=1}^{N} d_i T(\vec{v}_i)$ (where $c_{n+1} = \ldots = c_N = 0$).

$$\begin{aligned}
& g\left(\sum_{i=1}^{N} c_i T(\vec{v}_i) - \sum_{i=1}^{N} d_i T(\vec{v}_i), \sum_{i=1}^{N} c_i T(\vec{v}_i) - \sum_{i=1}^{N} d_i T(\vec{v}_i)\right) \\
= \; & g\left(\sum_{i=1}^{N} (c_i - d_i) T(\vec{v}_i), \sum_{j=1}^{N} (c_j - d_j) T(\vec{v}_j)\right) \\
= \; & \sum_{i=1}^{N} (c_i - d_i) \left(\sum_{j=1}^{N} \overline{(c_j - d_j)} g(T(\vec{v}_i), T(\vec{v}_j))\right) \\
= \; & \sum_{i=1}^{N} (c_i - d_i) \left(\sum_{j=1}^{N} \overline{(c_j - d_j)} g(\vec{v}_i, \vec{v}_j)\right) \\
= \; & g\left(\sum_{i=1}^{N} c_i \vec{v}_i - \sum_{i=1}^{N} d_i \vec{v}_i, \sum_{i=1}^{N} c_i \vec{v}_i - \sum_{i=1}^{N} d_i \vec{v}_i\right) \\
= \; & g(\vec{v} - \vec{v}, \vec{v} - \vec{v}) = 0.
\end{aligned}$$

The fact that H extends T follows from the definition of H, and the property that H preserves the inner product is easy to check:

$$\begin{aligned}
g(H(\vec{v}), H(\vec{w})) &= g\left(\sum_{i=1}^{n} c_i T(\vec{v}_i), \sum_{j=1}^{m} f_j T(\vec{v}_j)\right) \\
&= \sum_{i=1}^{n} c_i \left(\sum_{j=1}^{m} \overline{f_j} g(T(\vec{v}_i), T(\vec{v}_j))\right) \\
&= \sum_{i=1}^{n} c_i \left(\sum_{j=1}^{m} \overline{f_j} g(\vec{v}_i, \vec{v}_j)\right) = g\left(\sum_{i=1}^{n} c_i \vec{v}_i, \sum_{j=1}^{m} f_j \vec{v}_j\right) = g(\vec{v}, \vec{w}).
\end{aligned}$$

∎

References

[A] H. Anton, *Elementary Linear Algebra*, 10th ed., Wiley, New York, 2010.

[FIS] S. Friedberg, A. Insel, and L. Spence, *Linear Algebra*, 4th ed., Prentice-Hall, New Jersey, 2003.

[HK] K. Hoffman and R. Kunze, *Linear Algebra*, 2nd ed., Prentice-Hall, New Jersey, 1971.

[J] N. Jacobson, *Basic Algebra I*, Freeman & Co., San Francisco, 1974.