University of California, Berkeley

Literally everything I know about

Linear Algebra

Warren Alphonso

A very reductionist summary of Linear Algebra and its Applications by Lay, Lay, and McDonald, as well as Linear Algebra Done Wrong by Treil.

June 29, 2019

Contents

Chapter 1

Inner Product Spaces

Keep in mind that theory for inner product space is only developed for $\mathbb{R}$ and $\mathbb{C}$, so $\mathbb{F}$ will always denote one of those two fields in this chapter.

1.1 Inner Product

Definition 1.1.1. We define the **norm** of a vector to be the generalization of *length*. That is, the norm of a vector $x \in \mathbb{R}^n$ is

$$||x|| = \sqrt{x_1^2 + \cdots + x_n^2}$$

For any complex number $z = x + iy$, we can write $|z|^2 = x^2 + y^2 = z\overline{z}$, where $\overline{z}$ denotes the complex conjugate of z. So for any z in a complex field $\mathbb{C}^n$, we can write

$$z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} x_1 + iy_1 \\ \vdots \\ x_n + iy_n \end{bmatrix}$$

so it is natural to define the norm $||z||$ as

$$||z||^2 = \sum_{k=1}^{n} (x_k^2 + y_k^2) = \sum_{k=1}^{n} |z_k|^2$$

Definition 1.1.2. The **inner product** of two vectors $x, y \in \mathbb{R}^n$ is

$$(x, y) = x_1 y_1 + \cdots + x_n y_n = x^T y = y^T x$$

This yields another definition for the **norm**:

$$||x|| = \sqrt{(x, x)}$$

For complex fields, we need a definition of inner product such that $||z||^2 = (z, z)$. One definition that is consistent with this requirement will be our definition for the ***standard*** **inner product in** $\mathbb{C}^n$,

$$(z, w) = z_1 \overline{w_1} + \cdots + z_n \overline{w_n}$$

To simplify this, we will define the **Hermitian adjoint**, or simply **adjoint** A^*, by $A^* = \overline{A}^T$.

Using this, we can write

$$(z,w) = w^* z$$

The inner products we defined for $\mathbb{R}^n$ and $\mathbb{C}^n$ have the following properties:

1. Symmetry: $(x,y) = \overline{(y,x)}$
2. Linearity: $(\alpha x + \beta y, z) = \alpha(x,z) + \beta(y,z)$
3. Non-negativity: $(x,x) \geq 0$
4. Non-degeneracy: $(x,x) = 0$ if and only if $x = 0$

Note that properties 1 and 2 imply that

$$(x, \alpha y + \beta z) = \overline{(\alpha y + \beta z, x)} = \overline{\alpha}(x,y) + \overline{\beta}(x,z)$$

Lemma 1.1.1. *Let x be a vector in V. Then $x = 0$ if and only if*

$$(x,y) = 0 \qquad \forall y \in V$$

Proof. Since $(0,y) = 0$, we need to only show that $x = 0$ if $(x,y) = 0$. Subbing in $y = x$, we get $(x,x) = 0$ and property 3 asserts that $x = 0$.

Lemma 1.1.2. *Let x, y be vectors in V. Then $x = y$ if and only if*

$$(x,z) = (y,z) \qquad \forall z \in V$$

Proof. Using the above lemma, if we set $(x-y,z) = 0 \quad \forall z \in V$, then it follows that $x = y$ and $(x,z) = (y,z)$.

Theorem 1.1.3. *Suppose two operators $X, Y : A \to B$ satisfy*

$$(Ax,y) = (Bx,y) \qquad \forall x \in X, \forall y \in Y$$

Then $A = B$.

Proof. Using the previous lemma, we can fix x and take all $y \in Y$, which means $Ax = Bx$. Since this is true for all x, A and B are the same operator. ∎

Theorem 1.1.4 (Cauchy-Schwartz Inequality)**.**

$$|(x,y)| \leq ||x|| \cdot ||y||$$

Proof. If x or y is 0, then the proof is trivial. Assuming neither is 0, we will prove both the real and complex cases. But first consider only the real case:

$$0 \leq ||x - ty||^2 = (x - ty, x - ty) = ||x||^2 - 2t(x,y) + t^2||y||^2$$

Taking the derivative with respect to t and setting it to 0 gives us $t = \frac{(x,y)}{||y||^2}$. We will use this same t value for the following proof of the real and complex cases:

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

Using property 1 of inner products, we have

$$t = \frac{(x, y)}{\|y\|^2} = \frac{\overline{(y, x)}}{\|y\|^2}$$

Subbing in t, we get

$$0 \leq \|x\|^2 - \frac{|(xy)|^2}{\|y\|^2}$$

which completes the proof. ■

Theorem 1.1.5 (Triangle Inequality)**.**

$$\|x, y\| \leq \|x\| + \|y\|$$

Proof.

$$
\begin{aligned}
\|x + y\|^2 = (x + y, x + y) &= \|x\|^2 + \|y\|^2 + (x, y) + (y, x) \\
&\leq \|x\|^2 + \|y\|^2 + 2|(x, y)| \\
&\leq \|x\|^2 + \|y\|^2 + 2\|x\| \cdot \|y\| \\
&= (\|x\| + \|y\|)^2
\end{aligned}
$$

■

Theorem 1.1.6. *The following* ***polarization identities*** *allow us to construct the inner product from the norm: For $x, y \in \mathbb{R}^n$,*

$$(x, y) = \frac{1}{4}\left(\|x + y\|^2 - \|x - y\|^2\right)$$

For $x, y \in \mathbb{C}^n$,

$$(x, y) = \frac{1}{4}\left(\|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 - i\|x - iy\|^2\right)$$

Proof. For the real case,

$$
\begin{aligned}
\|x + y\|^2 - \|x - y\|^2 &= (x + y, x + y) - (x - y, x - y) \\
&= \|x\|^2 + \|y\|^2 + 2(x, y) - \|x\|^2 - \|y\|^2 + 2(x, y) \\
&= 4(x, y)
\end{aligned}
$$

For the complex case,

$$
\begin{aligned}
\sum_{k=0}^{3} i^k \|x + i^k y\|^2 &= \sum_{k=0}^{3} i^k (x + i^k y, x + i^k y) \\
&= \sum_{k=0}^{3} i^k \left(\|x\|^2 + \|y\|^2 + (x, i^k y) + (i^k y, x)\right) \\
&= \sum_{k=0}^{3} \left(i^k \|x\|^2 + i^k \|y\|^2 + (x, y) + (i^{2k} y, x)\right) \\
&= 4(x, y)
\end{aligned}
$$

where the last step follows from

$$\sum_{k=0}^{3} i^k = \sum_{k=0}^{3} i^{2k} = 0$$

■

Theorem 1.1.7 (Parallelogram Identity)**.** *Another important property of the norm is the parallelogram identity. For vectors u and v:*

$$||u+v||^2 + ||u-v||^2 = 2(||u||^2 + ||v||^2)$$

Proof. The theorem follows easily from the fact that the sum of the diagonals of a parallelogram equal the sum of all four sides. ■

To review, we have so far proved the following properties about the norm $||u||$:

1. Homogeneity: $||\alpha u|| = |\alpha| \cdot ||u||$
2. Triangle inequality: $||u+v|| \leq ||u|| + ||v||$
3. Non-negativity: $||u|| \geq 0$
4. Non-degeneracy: $||u|| = 0$ if and only if $u = 0$

In a vector space V, if we assign to each vector u a number $||u||$ that satisfies these 4 properties, we can say that the space V is a **normed space**.

1.2 Orthogonality

Definition 1.2.1. Two vectors u and v are **orthogonal**, denoted $u \perp v$, if and only if $(u,v) = 0$

Theorem 1.2.1. *If $u \perp v$, then*

$$||u+v||^2 = ||u||^2 + ||v||^2$$

Proof.

$$||u+v||^2 = ||u||^2 + ||v||^2 + (u,v) + (v,u) = ||u||^2 + ||v||^2$$

Since $(u,v) = (v,u) = 0$ because of orthogonality. ■

Definition 1.2.2. A vector u is **orthogonal to vector space** V if u is orthogonal to all vectors in V.

Theorem 1.2.2. *Let V be spanned by $v_1, \cdots, v_n$. Then $u \perp V$ if and only if*

$$u \perp v_k \qquad \forall k = 1, \cdots, n$$

Proof. Proving "only if" is trivial by the definition of $u \perp V$. Proving "if" comes easily after noticing that any vector can be rewritten as a linear combination of the basis vectors, so if u is perpendicular to all the basis vectors, then it is perpendicular to any other vector in V. ■

Definition 1.2.3. A set of vectors $v_1, \cdots, v_n$ are orthogonal if any two vectors in the set are orthogonal to each other. If $\|v_k\| = 1$ for all k, we call the set orthonormal.

Lemma 1.2.3 (Generalized Pythagorean Theorem). *Let $v_1, \cdots, v_n$ be an orthogonal system. Then*

$$\|\sum_{k=1}^{n} a_k v_k\|^2 = \sum_{k=1}^{n} |a_k|^2 \|v_k\|^2$$

Proof.

$$\|\sum_{k=1}^{n} a_k v_k\|^2 = \left(\sum_{k=1}^{n} a_k v_k, \sum_{j=1}^{n} a_j v_j\right) = \sum_{k=1}^{n} \sum_{j=1}^{n} a_k \overline{a_j} (v_k, v_j)$$

Since the set is orthogonal, (v_k, v_j) is only nonzero when $k = j$, so

$$= \sum_{k=1}^{n} |a_k|^2 \|v_k\|^2$$

Definition 1.2.4. An orthogonal set of vectors that is also a basis is called an **orthogonal basis**.

Typically, to find coordinates of a vector in a basis, we need to solve a system of equations. For orthogonal bases, it is much simpler. Suppose $v_1, \cdots, v_n$ is an orthogonal basis and let

$$x = \alpha_1 v_1 + \cdots + \alpha_n v_n$$

Taking the inner product with v_1 yields

$$(x, v_1) = (\sum_{j=1}^{n} \alpha_j (v_j, v_1) = \alpha_1 (v_1, v_1) = \alpha_1 \|v_1\|^2$$

Thus, to find any coordinate α_k of a vector x in orthogonal basis $v_1, \cdots, v_n$:

$$\alpha_k = \frac{(x, v_k)}{\|v_k\|^2}$$

This is a simple example of abstract orthogonal Fourier decomposition – simple because classical Fourier decomposition deals with infinite orthonormal systems.

1.3 Orthogonal Projection and Gram-Schmidt Orthogonalization

Definition 1.3.1. The **orthogonal projection** of a vector v onto the subspace E is the vector $w := P_E v$ such that $w \in E$ and $v - w \perp E$.

Theorem 1.3.1. *The orthogonal projection $w = P_E v$ minimizes the distance from v to E. In other words,*

$$\|v - w\| \leq \|v - x\| \qquad \forall x \in E$$

Additionally, if for some $x \in E$

$$\|v - w\| = \|v - x\|$$

then $x = w$.

Proof. Let $y = w - x \in E$. Then

$$v - x = v - w + w - x = v - w + y$$

Since $v - w \perp E$, we know $y \perp v - w$. By the Pythagorean Theorem,

$$\|v - x\|^2 = \|v - w\|^2 + \|y\|^2 \geq \|v - w\|^2$$

To finish the proof, note that equality only arises when $y = 0$, ie when $x = w$. ∎

There is a formula for finding an orthogonal projection if we know an orthogonal basis in E. Let $v_1, \cdots, v_n$ be an orthogonal basis in E. Then the projection $P_E v$ of a vector v is

$$P_E v = \sum_{k=1}^{n} a_k v_k \qquad where \qquad a_k = \frac{(v, v_k)}{\|v_k\|^2}$$

In other words,

$$P_E v = \sum_{k=1}^{n} \frac{(v, v_k)}{\|v_k\|^2} v_k$$

This is great if we have an orthogonal basis, but if even if we only have a basis in E, we can use the following algorithm to find an orthogonal basis.

Theorem 1.3.2 (Gram-Schmidt Orthogonalization Algorithm)**.** *Suppose we have linearly independent system $x_1, \cdots, x_n$. The Gram-Schmidt algorithm constructs from this an orthogonal system $v_1, \cdots, v_n$ such that*

$$span(x_1, \cdots, x_n) = span(v_1, \cdots, v_n)$$

Additionally, for all $r \leq n$

$$span(x_1, \cdots, x_r) = span(v_1, \cdots, v_r)$$

The algorithm is as follows:

1. *Define $v_1 := x_1$.*

   Define $E_1 := span(v_1) = span(x_1)$.

2. *Define $v_2 := x_2 - P_{E_1} x_2 = x_2 - \frac{(x_2, v_1)}{\|v_1\|^2} v_1$.*

   Define $E_2 := span(v_1, v_2) = span(x_1, x_2)$.

3. *Define $v_3 := x_3 - P_{E_2} x_3 = x_3 - \frac{(x_3, v_1)}{\|v_1\|^2} v_1 - \frac{(x_3, v_2)}{\|v_2\|^2} v_2$.*

   Define $E_3 := span(v_1, v_2, v_3) = span(x_1, x_2, x_3)$.

4. *Continue until we have n vectors and $span(v_1, \cdots, v_n) = span(x_1, \cdots, x_n)$. The formula for vector v_{r+1} given $v_1, \cdots, v_r$ is*

$$v_{r+1} := x_{r+1} - P_{E_r} x_{r+1} = x_{r+1} - \sum_{k=1}^{r} \frac{(x_{r+1}, v_k)}{\|v_k\|^2} v_k$$

Note that at each step, we are adding in x_{r+1} which means the resulting vector will not exist in E_r.

Proof. At each step, we add in x_{r+1} and then subtract its projection the subspace spanned by $x_1, \cdots, x_r$, meaning each additional vector is orthogonal to the ones previously defined. Since we set $v_1 = x_1$, we have proved the algorithm by induction. ∎

Since multiplication by a scalar does not change orthogonality, we can multiply vectors v_k returned by Gram-Schmidt by any non-zero numbers. One use case is to normalize the orthogonal vectors by dividing by their norms $\|v_k\|$ to yield an orthonormal system.

Definition 1.3.2. For a subspace E, its **orthogonal complement** $E^\perp$ is the set of all vectors orthogonal to E. Since at least 0 is orthogonal to E, $E^\perp$ is always a subspace.

By the definition of orthogonal projection, any vector in an inner product space V has a unique representation of the form

$$v = v_1 + v_2 \qquad v_1 \in E, v_2 \in E^\perp$$

This statement is usually written as $V = E \oplus E^\perp$.

Theorem 1.3.3. *For subspace E of V,*

$$(E^\perp)^\perp = E$$

Proof. We will show $E \subseteq (E^\perp)^\perp$ and $(E^\perp)^\perp \subseteq E$.
Let $u \in E$. Then $(u, v) = 0$ for all $v \in E^\perp$. Since u is orthogonal to every vector $v \in E^\perp$, then $u \in (E^\perp)^\perp$ so $E \subseteq (E^\perp)^\perp$.
Now let $u \in (E^\perp)^\perp$. Since $V = E \oplus E^\perp$, we can write $u = v + w$, where $v \in E$ and $w \in E^\perp$. This means that $u - v = w \in E^\perp$. Since we know $E \subseteq (E^\perp)^\perp$, we have $u \in (E^\perp)^\perp$ and $v \in (E^\perp)^\perp$, which means $u - v \in (E^\perp)^\perp$. Therefore, $u - v \in E^\perp \cap (E^\perp)^\perp$. Since the only vector that is orthogonal to itself is 0, $u = v$, and because $v \in E$, $(E^\perp)^\perp \subseteq E$. ∎

1.4 Least Square Solution

Recall that $Ax = b$ has a solution if and only if $b \in Range(A)$. In real life, it is impossible to avoid errors. The simplest way to approximate a solution is to choose an approximation $\hat{x}$ to minimize the error $e = \|A\hat{x} - b\|$. This is the **least square solution**.
We know $A\hat{x}$ is the orthogonal projection $P_{Range(A)}b$ if and only if $b - A\hat{x} \perp Range(A)$. Using the column space interpretation of range, this is equivalent to

$$b - A\hat{x} \perp a_k \qquad \forall k = 1, \cdots, n$$

That means

$$0 = (b - A\hat{x}, a_k) = a_k^*(b - A\hat{x}) \qquad \forall k = 1, \cdots, n$$

We can join the rows a_k^* together to get

$$A^*(b - A\hat{x}) = 0$$

which is equivalent to the **normal equation**

$$A^*A\hat{x} = A^*b$$

The solution $\hat{x}$ to this equation grants us the least square solution of $A\hat{x} = b$. This makes it easy to notice that the least square solution is unique if and only if A^*A is invertible.
If $\hat{x}$ is the solution to the normal equation, then $A\hat{x} = P_{Range(A)}b$. So in order to find the actual projection of b onto $Range(A)$, we need to solve the normal equation and then multiply the solution by A. Formally,

$$P_{Range(A)}b = A(A^*A)^{-1}A^*b$$

Because this is true for all b, the formula for the matrix of the orthogonal projection onto $Range(A)$ is

$$P_{Range(A)} = A(A^*A)^{-1}A^*$$

Theorem 1.4.1. *For an $m \times n$ matrix A*

$$Ker(A) = Ker(A^*A)$$

Recall Kernel is equivalent to Null Space.

Proof. We will show $Ker(A) \subseteq Ker(A^*A)$ and $Ker(A^*A) \subseteq Ker(A)$.
To prove the latter, suppose we have a vector $u \in Ker(A)$ so that $Au = 0$. Then $A^*Au = A^*(Au) = A^*0 = 0$, which means $u \in Ker(A^*A)$.
To prove the former, suppose we have a vector $v \in Ker(A^*A)$. We want to show that $Av = 0$. One way of doing so is to show that its norm is 0.

$$\|Av\|^2 = (Av, Av) = (A^*v^*, A^*v^*) = A^*(v^*, A^*v^*) = A^*(Av, v) = (A^*Av, v) = (0, v) = 0$$

■

1.5 Adjoint of a Linear Transformation

Recall that the *Hermitian adjoint* A^* of matrix A is defined as the complex conjugate of each entry in A^T.

Theorem 1.5.1.

$$(Ax, y) = (x, A^*y) \qquad \forall x \in \mathbb{C}^n, \forall y \in \mathbb{C}^m$$

Proof.

$$(Ax, y) = y^*Ax = (A^*y)^*x = (x, A^*y)$$

The second equality uses the fact that because the adjoint consists of a transpose, we have $(AB)^* = B^*A^*$ and $(A^*)^* = A$. ■

This identity is used to define the adjoint operator.

Lemma 1.5.2. *The adjoint is unique.*

Proof. Suppose B satisfies $(Ax, y) = (x, By) \quad \forall x, y$, then we can write

$$(Ax, y) = (x, A^*y) = (x, By)$$

which means $A^* = B$.

Properties of the adjoin operators (matrices):

1. $(A + B)^* = A^* + B^*$
2. $(\alpha A)^* = \overline{\alpha} A^*$
3. $(AB)^* = B^*A^*$
4. $(A^*)^* = A$
5. $(y, Ax) = (A^*y, x)$

Theorem 1.5.3 (Relation between fundamental subspaces). *Let $A : V \to W$ be an operator acting from one inner product space to another. Then*

1. $Ker(A^*) = (Range(A))^\perp$
2. $Ker(A) = (Range(A^*))^\perp$
3. $Range(A) = (Ker(A^*))^\perp$
4. $Range(A^*) = (Ker(A))^\perp$

Note that earlier we defined the fundamental subspaces using A^T instead of A^* because when discussing only $\mathbb{R}$ there was no difference.

Proof. Note that statements 1/3 and 2/4 are equivalent because for any subspace E, we have $(E^\perp)^\perp = E$. Also note that statement 2 is exactly statement 1 applied to the operator A^* since $(A^*)^* = A$.
Thus we only need to prove statement 1.
A vector $x \in (Range(A))^\perp$ means that x is orthogonal to all vectors of the form Ay, that is

$$(x, Ay) = 0 \qquad \forall y$$

Since $(x, Ay) = (A^*x, y)$, this is equivalent to

$$(A^*x, y) = 0 \qquad \forall y$$

This means that $A^*x = 0$, which means $x \in Ker(A^*)$. ∎

1.6 Isometries and Unitary Operators

Definition 1.6.1. An operator $U : X \to Y$ is called an **isometry** if it preserves the norm,

$$\|Ux\| = \|x\| \qquad \forall x \in X$$

Theorem 1.6.1. *An operator $U : X \to Y$ is an isometry if and only if it preserves the inner product, ie if and only if*

$$(x, y) = (Ux, Uy) \qquad \forall x, y \in X$$

Proof. We use the polarization identities previously described. If X is a complex space

$$\begin{aligned}
(Ux, Uy) &= \frac{1}{4} \sum_{\alpha = \pm 1, \pm i} \alpha \|Ux + \alpha Uy\|^2 \\
&= \frac{1}{4} \sum_{\alpha = \pm 1, \pm i} \alpha \|U(x + \alpha y)\|^2 \\
&= \frac{1}{4} \sum_{\alpha = \pm 1, \pm i} \alpha \|x + \alpha y\|^2 = (x, y)
\end{aligned}$$

If X is a real space

$$\begin{aligned}(Ux, Uy) &= \frac{1}{4}(||Ux + Uy||^2 - ||Ux - Uy||^2) \\ &= \frac{1}{4}(||U(x+y)||^2 - ||U(x-y)||^2) \\ &= \frac{1}{4}(||x+y||^2 - ||x-y||^2) = (x, y)\end{aligned}$$

■

Lemma 1.6.2. *An operator $U : X \to Y$ is an isometry if and only if $U^*U = I$.*

Proof. If $U^*U = I$, then

$$(x, x) = (U^*Ux, x) = (Ux, Ux) \qquad \forall x \in X$$

Since $||x|| = ||Ux||$, U is an isometry.
If U is an isometry, then by the above theorem and definition of adjoint

$$(U^*Ux, y) = (Ux, Uy) = (x, y) \qquad \forall x, y \in X$$

which means $U^*U = I$.

This lemma implies that an isometry is always left invertible since $U^*U = I$.

Definition 1.6.2. An isometry $U : X \to Y$ is called a **unitary operator** if it is invertible.

Lemma 1.6.3. *An isometry $U : X \to Y$ is a unitary operator if and only if $dim(X) = dim(Y)$.*

Proof. If $dim(X) = dim(Y)$, then U is square. Since we know U is left invertible, it must also then be invertible.
If U is unitary, it is invertible, so $dim(X) = dim(Y)$ since only square matrices are invertible.

Properties of unitary operators that follow from our proofs:

1. $U^{-1} = U^*$
2. If U is unitary, $U^* = U^{-1}$ is also unitary.
3. If U is an isometry and $v_1, \cdots, v_n$ is an orthonormal basis, then $Uv_1, \cdots, Uv_n$ is an orthonormal basis.
4. The product of unitary operators is a unitary operator.

Lemma 1.6.4.

$$det(A^*) = \overline{det(A)}$$

Proof. Recall that the determinant of a matrix is equal to the product of its eigenvalues. We will show that the for any eigenvalue λ of A, $\overline{\lambda}$ is an eigenvalue of A^*.
Note that λ is **not** an eigenvalue of A if and only if $A - \lambda I$ is invertible, which happens if and only if there exists an operator B such that

$$B(A - \lambda I) = (A - \lambda I)B = I$$

Taking the adjoints of all three sides means the above is equivalent to

$$(A^* - \overline{\lambda} I)B^* = B^*(A^* - \overline{\lambda} I) = I$$

Thus $A - \lambda I$ is invertible if and only if $A^* - \overline{\lambda} I$ is invertible, which means if λ is an eigenvalue of A, $\overline{\lambda}$ is an eigenvalue of A^*.

Theorem 1.6.5. *If U is a unitary matrix, then*

$$det(U) = \pm 1$$

If λ is an eigenvalue of U, then

$$\lambda = \pm 1$$

Proof. Let $det(U) = z$. Since $det(U^*) = \overline{det(U)}$, we have

$$|z|^2 = \overline{z}z = det(U^*U) = det(I) = 1$$

To prove statement 2, notice that if $Ux = \lambda x$, then

$$||Ux|| = ||\lambda x|| = |\lambda| \cdot ||x||$$

which means $|\lambda| = 1$ since $||Ux|| = ||x||$. ■

Definition 1.6.3. Operators A and B are called **unitarily equivalent** if there exists a unitary operator U such that $A = UBU^*$. Since for any unitary U, we have $U^{-1} = U^*$, any two unitarily equivalent matrices are similar as well.
The converse is **not** true.

The following theorem gives a way to construct a counter example to prove similar matrices are not always unitarily equivalent.

Theorem 1.6.6. *A matrix A is unitarily equivalent to a diagonal one if and only if it has an orthogonal (orthonormal) basis of eigenvectors.*

Proof. Using diagonalization, we can write $A = UBU^*$ and let $Bx = \lambda x$. Then $AUx = UBx = \lambda Ux$, which means Ux is an eigenvector of A.
Only if: Let A be unitarily equivalent to a diagonal matrix D, ie $A = UDU^*$. Because D is diagonal, the vectors e_k of the standard basis are eigenvectors of D, so Ue_k are eigenvectors of A. Since U is unitary, $Ue_1, \cdots, Ue_n$ is an orthonormal basis.
If: Let A have an orthogonal basis $u_1, \cdots, u_n$ of eigenvectors. By dividing each vector by its norm, we can assure we have an orthonormal basis. By letting D be the matrix A in the basis $u_1, \cdots, u_n$, we know D will be a diagonal matrix.
By setting U to be the matrix with columns $u_1, \cdots, u_n$, we know U is unitary since its columns form an orthonormal basis (orthogonality implies invertibility and normality implies norm preservation). The change of coordinate formula implies

$$A = [A]_{SS} = [I]_{SB}[A]_{BB}[I]_{BS} = UDU^{-1} = UDU^*$$

where the last step follows from $U^{-1} = U^*$ for unitary matrices. ■