Contents

Chapter 1

Linear Systems

Talk about linear systems with a view to elementary operations.

1. Row operations

2. Rank of a matrix

3. Full rank equiv to invertible equiv to unique solution (nonsingular)

4. Invertible is row equivalent to identity (equivalence relation)

Chapter 2

Special Matrices

1. Unitary matrices

Chapter 3

Basis

1. Definition of basis: LI spanning set. Consequences: linear independence implies unique basis expansion of any vector. How to find? Solve consistent linear system. If basis orthogonal (Gram-schmidt leaves no excuse not to be), then the expansion is easy to compute using inner products.

2. Alt defns: minimal spanning set, maximal LI subset. These are all equivalent (show).

3. Prove that every vector space has one!

4. A choice of basis allows us to find a matrix representation for any linear transformation. That is, given a linear transformation and a basis, we can represent the transformation as a matrix multiplication.

3.1   Definition of Basis, Consequences

A *basis* for a vector subspace $V \subset \mathbb{R}^n$ is a linearly independent spanning set $\mathcal{B}$ for V. That is, it is a set of vectors

$$\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\} \subset V$$

such that any vector $\vec{x} \in V$ can be written as a (finite) linear combination

$$\vec{x} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_k\vec{v}_k$$

of elements in $\mathcal{B}$. The linear independence property of the basis makes this basis expansion unique: if

$$\vec{x} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_k\vec{v}_k$$

and

$$\vec{x} = d_1\vec{v}_1 + d_2\vec{v}_2 + \cdots + d_k\vec{v}_k$$

then we have

$$\begin{aligned}
0 &= \vec{x} - \vec{x} \\
&= c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_k\vec{v}_k - d_1\vec{v}_1 - d_2\vec{v}_2 - \cdots - d_k\vec{v}_k \\
&= (c_1 - d_1)\vec{v}_1 + (c_2 - d_2)\vec{v}_2 + \cdots + (c_k - d_k)\vec{v}_k
\end{aligned}$$

and so by linear independence of $\mathcal{B}$ we must have $c_i = d_i$ for all $i = 1, 2, \ldots, k$. Thus, the basis expansion of $\vec{x}$ is unique. Thus, given a basis for a vector space, we have a unique basis expansion of any vector in the space. In fact, this uniqueness is satisfied by any vector expansion in terms of a linearly independent set; the spanning condition requires that such a unique expansion exists for every vector in the space.

How can we actually compute this basis expansion? If we know that

$$\vec{x} = c_1 \vec{v}_1 + c_2 \vec{v}_2 + \cdots + c_k \vec{v}_k$$

for some unknown scalars $c_1, c_2, \ldots, c_k$, then this can be rewritten as

$$\vec{x} = \begin{bmatrix} | & | & & | \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_k \\ | & | & & | \end{bmatrix} \vec{c}$$

where $\vec{c}^T = (c_1, c_2, \ldots, c_k)$. This is a linear system, and is solved in the usual way for the coefficients $\vec{c}$. (SHOW THAT THIS IS CONSISTENT)

In the case where the basis $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is orthogonal (the basis vectors are pairwise orthogonal), this expansion is much more easily computed.

Prop 3.1. *Given an orthogonal basis $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ for subspace $V \subset \mathbb{R}^n$, we have for any $\vec{x} \in V$ the expansion*

$$\vec{x} = \sum_{i=1}^{k} \frac{\langle \vec{x}, \vec{v}_i \rangle^2}{\|\vec{v}_i\|} \vec{v}_i$$

(PF) Since $\mathcal{B}$ is a basis, we have some expansion

$$\vec{x} = \sum_{i=1}^{k} c_i \vec{v}_i$$

Taking the inner product with $\vec{v}_i$, we have

$$\langle \vec{x}, \vec{v}_j \rangle = \sum_{i=1}^{k} c_i \langle \vec{v}_i, \vec{v}_j \rangle$$
$$= c_j \langle \vec{v}_j, \vec{v}_j \rangle$$
$$= c_j \|\vec{v}_j\|^2$$

so that

$$c_j = \frac{\langle \vec{x}, \vec{v}_j \rangle}{\|\vec{v}_j\|^2}$$

and we have the basis expansion.

Conversely, if this expansion holds for any vector in V, then it holds in particular for the basis vectors. For $\vec{v}_j$, then, we have

$$\vec{v}_j = \sum_{i=1}^{k} \frac{\langle \vec{v}_j, \vec{v}_i \rangle}{\|\vec{v}_i\|^2} \vec{v}_i$$

5

Since this expansion is unique (due to linear independence of basis $\mathcal{B}$), from the left-hand side we can see that the only nonzero component of this expansion on the right-hand side is

$$\frac{\langle \vec{v}_j, \vec{v}_j \rangle}{\|\vec{v}_j\|^2} = 1$$

and $\langle \vec{v}_i, \vec{v}_p \rangle = 0$ for $i \neq p$. Thus, the basis is orthogonal.

Note that the popular Gram-Schmidt process allows us to orthogonalize any set of vectors. Hence, we can generally assume that our bases are orthogonal (or even othornormal) since these greatly simplify our calculations.

3.1.1 Alternate Characterizations

Given a finite-dimensional subpace $V \subset \mathbb{R}^n$, a basis is a linear independent spanning set. It turns out that there are alternate and equivalent characterizations of a vector space basis:

Prop 3.2. *Let $V \subset \mathbb{R}^n$ be a k-dimensional vector space with subset $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$. Then the following are equivalent:*

1. *$\mathcal{B}$ is a basis for V (a linearly independent spanning set).*

2. *$\mathcal{B}$ is a maximal linearly independent subset of V.*

3. *$\mathcal{B}$ is a minimal spanning set for V.*

Proof. First, assume $\mathcal{B}$ is a basis for V. Let $\vec{v} \in V \setminus \mathcal{B}$ be an arbitrary nonzero, non-basis vector. Then since the basis spans V, we have constants c_i, $i = 1, \ldots, k$ such that

$$\vec{v} = c_1 \vec{v}_1 + c_2 \vec{v}_2 + \cdots + c_k \vec{v}_k$$

or, upon rearranging,

$$c_1 \vec{v}_1 + c_2 \vec{v}_2 + \cdots + c_k \vec{v}_k - \vec{v} = 0$$

which is a nontrivial linear combination of the vectors $\mathcal{B} \cup \{\vec{v}\}$ which is equal to the zero vector; thus, $\mathcal{B} \cup \{\vec{v}\}$ is linearly dependent. Thus, $\mathcal{B}$ is a maximal linearly independent subset of V.

Next, if $\mathcal{B}$ is still a basis for V, then assume $\mathcal{B}' \subset \mathcal{B}$ is a subset with $V = \text{Span}(\mathcal{B}')$. Then since $\mathcal{B}'$ is a proper subset of $\mathcal{B}$, there is some vector $\vec{v}_m \in \mathcal{B} \setminus \mathcal{B}'$. Since $\vec{v}_m \in V$, there is an expansion

$$\vec{v}_m = c_1 \vec{v}_1 + c_2 \vec{v}_2 + \cdots + c_{m-1} \vec{v}_{m-1} + c_{m+1} \vec{v}_{m+1} + \cdots + c_k \vec{v}_k$$

in the basis $\mathcal{B}'$ which does not contain $\vec{v}_m$ (and perhaps other vectors in $\mathcal{B}$). Then we can rearrange to obtain

$$c_1 \vec{v}_1 + c_2 \vec{v}_2 + \cdots + c_{m-1} \vec{v}_{m-1} - \vec{v}_m + c_{m+1} \vec{v}_{m+1} + \cdots + c_k \vec{v}_k = 0$$

which is a nontrivial linear combination of vectors of $\mathcal{B}$ (maybe not all of them) which equals the zero vector: $\mathcal{B}$ is a linearly dependent set. This conclusion contradicts our hypothesis, and so we have basis $\mathcal{B}$ a minimal spanning set for V.

Conversely, assume $\mathcal{B}$ is a maximal linearly independent subset of V. It remains to show that $\mathcal{B}$ spans V. Let $\vec{v} \in V$ be an arbitrary vector. If $\vec{v} \in \mathcal{B}$, then certainly $\vec{v} \in \operatorname{Span}(\mathcal{B})$. On the other hand, if $\vec{v} \notin \operatorname{Span}(\mathcal{B})$, then $\mathcal{B} \cup \{\vec{v}\}$ is linearly dependent, i.e., there are constants $c_1, c_2, \ldots, c_k, c$ not all zero such that

$$c_1 \vec{v}_1 + c_2 \vec{v}_2 + \cdots + c_k \vec{v}_k + c\vec{v} = 0$$

Clearly $c \neq 0$ since otherwise we would have

$$c_1 \vec{v}_1 + c_2 \vec{v}_2 + \cdots + c_k \vec{v}_k = 0$$

with not all coefficients zero, and so $\mathcal{B}$ would be linearly dependent. Instead, since $c \neq 0$, we can rearrange and divide to obtain

$$\vec{v} = \frac{c_1}{c} \vec{v}_1 + \frac{c_2}{c} \vec{v}_2 + \cdots + \frac{c_k}{c} \vec{v}_k$$

so that $\vec{v} \in \operatorname{Span}(\mathcal{B})$. Thus, $\mathcal{B}$ spans V, and so $\mathcal{B}$ is a basis for V.

Last, assume that $\mathcal{B}$ is a minimal spanning set for V. We must show that it is linearly independent. Assume by way of contradiction that there is a nontrivial linear combination

$$c_1 \vec{v}_1 + c_2 \vec{v}_2 + \cdots + c_k \vec{v}_k = 0$$

Then assume without loss of generality that $c_1 \neq 0$; upon rearranging, we obtain

$$\vec{v}_1 = -\frac{c_2}{c_1} \vec{v}_2 - \frac{c_3}{c_1} \vec{v}_2 - \cdots - \frac{c_k}{c_1} \vec{v}_k$$

Then we have $\vec{v}_1$ as a linear combination of the vectors $\mathcal{B} \setminus \{\vec{v}_1\}$, and so any linear combination of the vectors in $\mathcal{B}$ can be rewritten as a linear combination of those vectors without $\vec{v}_1$. Since $\mathcal{B}$ spans V, so too does $\mathcal{B} \setminus \{\vec{v}_1\}$, contradicting our assumption of $\mathcal{B}$ as a minimal spanning set. ∎

3.1.2 Existence

3.2 Isomorphism from Abstract V to $\mathbb{R}^n$

3.2.1 Matrix Representations

This is not quite right. if V is an abstract vector space, then there is an isomorphism

$$\vec{x} = x_1 \vec{v}_1 + \cdots x_k \vec{v}_k \mapsto \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}$$

from V to $\mathbb{R}^k$. In $V = \mathbb{R}^k$, it is easy to see that this representation works since the vectors themselves already have an interpretation as tuples, but in abstract vector spaces this construction still works.

We can discuss vector spaces in the abstract as simply sets of objects (along with an underlying field) upon which a vector addition and scalar multiplication are defined which satisfy the usual axioms. If we want to do any sort of practical computations, though, we need to express our vectors in a basis. If we have an abstract vector $\vec{x} \in V$ and a basis $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ for V, then we can write

$$\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k$$

for some scalars $x_1, x_2, \ldots, x_k$. Then we say the scalars x_i are the coordinates of $\vec{x}$ with respect to the ordered basis $\mathcal{B}$. Note that we can write this linear combination as a matrix product

$$\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k = \begin{bmatrix} | & | & & | \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_k \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}$$

where the matrix on the right is the only part of the product which depends on the scalars x_i which uniquely identify $\vec{x}$. Thus, we can identify a vector $\vec{x}$ with its matrix of coordinates in the ordered basis $\mathcal{B}$ and call this the representation of $\vec{x}$ in the $\mathcal{B}$ basis, written

$$[\vec{x}]_{\mathcal{B}} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}$$

or more simply as $\vec{x}$ when the basis is assumed to be known.

In fact, the choice of a basis allows us to express linear transformations and mappings in a standard way. For example, let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be a linear transformation. Then given any arbitrary $\vec{x} \in \mathbb{R}^3$, we have a basis expansion

$$\vec{x} = x_1\hat{e}_1 + x_2\hat{e}_2 + x_3\hat{e}_3$$

Then by linearity of T, we have

$$T(\vec{x}) = x_1 T(\hat{e}_1) + x_2 T(\hat{e}_2) + x_3 T(\hat{e}_3)$$

so that the action of T on any vector in $\mathbb{R}^3$ is determined entirely by its action on the basis vectors. Note that the quantities $T(\hat{e}_i)$ are themselves vectors in $\mathbb{R}^3$, and can itself be expanded in the basis $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$:

$$T(\hat{e}_i) = a_{1i}\hat{e}_1 + a_{2i}\hat{e}_2 + a_{3i}\hat{e}_3$$

Substituting these expansions in the previous expression, we can write

$$T(\vec{x}) = (x_1 a_{11} + x_2 a_{12} + x_3 a_{13})\,\hat{e}_1 + (x_1 a_{21} + x_2 a_{22} + x_3 a_{23})\,\hat{e}_2 + (x_1 a_{31} + x_2 a_{32} + x_3 a_{33})\,\hat{e}_3$$

which is a vector in $\mathbb{R}^3$. The matrix representation of this vector is

$$T(\vec{x}) = \begin{bmatrix} x_1 a_{11} + x_2 a_{12} + x_3 a_{13} \\ x_1 a_{21} + x_2 a_{22} + x_3 a_{23} \\ x_1 a_{31} + x_2 a_{32} + x_3 a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = A\vec{x}$$

and so the action of T on any vector in $\mathbb{R}^3$ can be represented by a matrix multiplication. That is, given a basis on a vector space, linear transformations of that space can always be written in a matrix representation, through which their action on vectors is expressed as matrix multiplication. Conversely, any (square) matrix transformation of a vector is a linear transformation. In the case of linear transformations, where the domain and codomain are the same space, the bases are the same and so the matrix representation is square.

More generally, we state the conclusion as a proposition:

Prop 3.3. *Given vector spaces $(V, \mathcal{B}_V)$, $(W, \mathcal{B}_W)$, any linear mapping $T : V \to W$ has a matrix representation. That is, we have $T(\vec{x}) = A\vec{x}$ for every $\vec{x} \in V$ for a suitable choice of A.*

Proof. We will prove the statement by constructing the matrix representation; this is a relatively straightforward generalization of the example above. Assume that V and W are vector spaces with bases $\mathcal{B}_V = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ and $\mathcal{B}_W = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$, respectively. Let $\vec{x} \in V$ be arbitrary with basis expansion

$$\vec{x} = x_1 \vec{v}_1 + x_2 \vec{v}_2 + \cdots + \vec{v}_k$$

Then since T is linear, we have

$$T(\vec{x}) = x_1 T(\vec{v}_1) + x_2 T(\vec{v}_2) + \cdots + x_k T(\vec{v}_k)$$

Note that this vector $T(\vec{x})$ resides in W, the codomain of T, and we have written it as a linear combination of the images of the vectors in $\mathcal{B}_V$, $T(\vec{v}_i)$. Since $\mathcal{B}_W$ spans W, we can write each $T(\vec{v}_i)$ in the basis for W. That is, we have

$$T(\vec{v}_1) = a_{11} \vec{w}_1 + a_{21} \vec{w}_2 + \cdots + a_{m1} \vec{w}_m$$
$$T(\vec{v}_2) = a_{12} \vec{w}_1 + a_{22} \vec{w}_2 + \cdots + a_{m2} \vec{w}_m$$
$$\vdots$$
$$T(\vec{v}_i) = a_{1i} \vec{w}_1 + a_{2i} \vec{w}_2 + \cdots + a_{mi} \vec{w}_m$$
$$\vdots$$
$$T(\vec{v}_k) = a_{1k} \vec{w}_1 + a_{2k} \vec{w}_2 + \cdots + a_{mk} \vec{w}_m$$

Note that we can now write

$$[T(\vec{v}_i)]_{\mathcal{B}_W} = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{bmatrix}$$

so that in the basis $\mathcal{B}_W$ we have

$$[T(\vec{x})]_{\mathcal{B}_W} = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_k \begin{bmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{mk} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ & & & \\ a_{m1} & a_{m2} & \cdots & a_{mk} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}$$

and so we have represented an arbitrary linear map $T : V \to W$ as a matrix multiplication:

$$[T(\vec{x})]_{\mathcal{B}_W} = A[\vec{x}]_{\mathcal{B}_V}$$

Note that in the case where $V = W$, the mapping T is a linear transformation on V and the matrix representation is square. ∎

Compare this proof with the change-of-basis formulas in section 5.1. In this situation, our mapping sends us from one vector space into another, and the two spaces generally have different bases. In the change of basis discussion, we are merely changing the basis of a vector. We will discuss in future sections the relationships between these concepts, e.g., how a change of basis is a linear transformation.

In fact, we can even go a step further and say that the space of linear transformations is isomorphic to the set of real matrices (of appropriate shape)!

Chapter 4

Dimension

Clear up all of the confusion and theorems around dimension of a vector (sub)space.

1. Show that it is uniquely defined.

2. Relation to rank of matrix with basis vectors as columns.

3. Why does this matrix have full rank?

Chapter 5

Change of Basis

What does a basis really mean? Finite-dimensional case only.

1. A vector itself never changes; it is always the same object.

2. We need a basis to be able to describe a vector concretely as a list of coordinates. A different basis requires different coordinates. Coords tell us "how much" of each basis vector is used to build our vector.

3. When we have a new basis, we generally have the new basis vectors written in terms of the old ones. We can use this relation between the bases to find a matrix which transforms coordinates (a matrix representation) from one basis to another. Do this.

4. If we can change vectors between bases, then we can also change our representations of arbitrary linear transformations: change the vector, apply the transformation, and change back (draw diagram).

5.1 Vector Change of Basis

Given a vector space V and a basis $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ for V, we know that we can write $\vec{x}$ in a basis expansion

$$\vec{x} = x_1 \vec{v}_1 + x_2 \vec{v}_2 + \cdots + x_k \vec{v}_k$$

and in the equivalent matrix representation

$$[\vec{x}]_\mathcal{B} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}$$

Given a new basis $\mathcal{C} = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$ for V, we would like to express $\vec{x}$ in the basis $\mathcal{C}$. In general, we are given the new basis vectors $\vec{w}_j$ in terms of the old basis vectors $\vec{v}_i$:

$$\vec{w}_j = a_{1j} \vec{v}_1 + a_{2j} \vec{v}_2 + \cdots a_{kj} \vec{v}_k = \sum_{i=1}^{k} a_{ij} \vec{v}_i$$

Such an expression is guaranteed to exist, since $\mathcal{B}$ is a basis for V; further, we require this expansion to be known to us, i.e., we know the coefficients/coordinates a_{ij}. Otherwise, there is no real we that we can relate the two bases. It also makes sense to know this, since our desire to change to another basis implies that we might have some knowledge about this basis, i.e., how it relates to the old basis. We can write this in the matrix form as

$$[\vec{w}_j]_{\mathcal{B}} = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{kj} \end{bmatrix} \tag{5.1}$$

and this tells us how much of each basis vector in $\mathcal{B}$ goes into each basis vector in $\mathcal{C}$.

Since $\mathcal{C}$ is a basis for V, we have an (unique) expansion of $\vec{x}$ as

$$\vec{x} = x_1' \vec{w}_1 + x_2' \vec{w}_2 + \cdots + x_k' \vec{w}_k \tag{5.2}$$

The goal of our change of basis is to determine the coordinates x_i', since we seek

$$[\vec{x}]_{\mathcal{C}} = \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_k' \end{bmatrix} \tag{5.3}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} = \begin{bmatrix} | & | & & | \\ [\vec{w}_1]_{\mathcal{B}} & [\vec{w}_2]_{\mathcal{B}} & \cdots & [\vec{w}_k]_{\mathcal{B}} \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_k' \end{bmatrix} \tag{5.4}$$

or, more concisely,

$$[\vec{x}]_{\mathcal{B}} = P[\vec{x}]_{\mathcal{C}}$$

where P is the matrix whose columns are the ordered basis vectors of W in $\mathcal{C}$ expressed in the basis $\mathcal{B}$ of V. This expression is fundamental and very intuitive, since the coordinates of $[\vec{x}]_{\mathcal{C}}$ tell how much of each vector in $\mathcal{C}$ is used to reconstruct $\vec{x}$. The matrix multiplication with P converts these "amounts of vectors in $\mathcal{C}$" into the corresponding amounts of each vector in $\mathcal{B}$ from which each vector in $\mathcal{C}$ is comprised. This is the easy device used to remember this behavior. This expression is straightforward to compute, but takes some time; it is easier to try to reason through it.

Note that to accomplish the original goal, to map $[\vec{x}]_{\mathcal{B}} \mapsto [\vec{x}]_{\mathcal{C}}$ can be achieved as

$$[\vec{x}]_{\mathcal{C}} = P^{-1}[\vec{x}]_{\mathcal{B}}$$

provided this inverse exists!

Does it always exist? The matrix P is comprised of columns which are the vectors in $\mathcal{C}$. Since $\mathcal{C}$ is a basis, these columns are linearly independent and so P has full rank and is thus invertible.

In the special case where $\mathcal{B}$ and $\mathcal{C}$ are both orthogonal bases, note that $P^T P = I_k$ and so $P^T = P^{-1}$, i.e., P is an orthogonal matrix. This is a little confusing: if the basis vectors

in $\mathcal{C}$ are orthogonal, and columns of P are these vectors in another basis, shouldn't those columns be orthogonal? That is, shouldn't the change of basis of these basis vectors preserve orthogonality? As it turns out, this is not true. Take the example

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$$

and

$$\mathcal{C} = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

Both $\mathcal{B}$ and $\mathcal{C}$ are basis for $\mathbb{R}^2$, but $\mathcal{C}$ is orthogonal while $\mathcal{B}$ is decidedly not (in the standard inner product). For arbitrary $\vec{x} \in \mathbb{R}^2$, we have

$$[\vec{x}]_\mathcal{B} = P[\vec{x}]_\mathcal{C}$$

where P is the change-of-basis matrix

$$P = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}$$

This is easily verified: the columns of P are the basis vectors in $\mathcal{C}$ written in the basis $\mathcal{B}$.

The basis vectors in $\mathcal{C}$ are orthogonal, but not when expressed in the basis $\mathcal{B}$.

Note that the columns of an orthogonal matrix are ortho*normal* in the usual dot product; they are not only pairwise orthogonal, but they are unit vectors. Also, the IP itself is a fixed function. It is independent of basis, but it is the *formula* which changes

5.2 Change of basis: endomorphism

Consider a linear transformation $T : V \to V$ acting on vector space V. Recall that given a basis

$$\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$$

for V, we can express T as a matrix multiplication, i.e.,

$$T(\vec{x}) = A[\vec{x}]_\mathcal{B}$$

where the columns of A are the image of the basis vectors under T:

$$A = \begin{bmatrix} | & | & & | \\ T(\vec{v}_1) & T(\vec{v}_2) & \cdots & T(\vec{v}_k) \\ | & | & & | \end{bmatrix}$$

and each $T(\vec{v}_i)$ is written in the basis $\mathcal{B}$ of V.

Assume now that we want to change our basis. In order for the operation of T to not change when we switch basis — it shouldn't — we must adapt our basis-based representation

of the matrix T. That is, we must come up with some matrix A' which will represent the linear map T in the new basis. Let the new basis be

$$\mathcal{C} = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$$

Then we have the change-of-basis matrix P which relates the coordinate representations of any vector $\vec{x}$ as

$$[\vec{x}]_{\mathcal{B}} = P[\vec{x}]_{\mathcal{C}}$$

Recall that the columns of P are the basis vectors in $\mathcal{C}$ written in the basis $\mathcal{B}$, and P necessarily has full rank. Then note that $T(\vec{x})$ is itself a vector in V, and so is expressed as a coordinate vector in either the basis $\mathcal{B}$ or $\mathcal{C}$. These coordinate vectors, as vectors in V, are related by the same change-of-basis matrix P:

$$[T(\vec{x})]_{\mathcal{B}} = P[T(\vec{x})]_{\mathcal{C}}$$

It is important to note that the function T and its value on $\vec{x}$ are not changing, merely their matrix representation. We then have

$$[T(\vec{x})]_{\mathcal{B}} = A[\vec{x}]_{\mathcal{B}} = AP[\vec{x}]_{\mathcal{C}}$$

but also

$$[T(\vec{x})]_{\mathcal{B}} = P[T(\vec{x})]_{\mathcal{C}} = PA'[\vec{x}]_{\mathcal{C}}$$

and so we must have the relationship

$$AP = PA'$$

or

$$A' = P^{-1}AP$$

relating the matrix representation A of T in the $\mathcal{B}$ basis to the representation A' in the $\mathcal{C}$ basis.

This is easily remembered by thinking about the matrix multiplication on the right-hand side as function composition. First, the vector in the $\mathcal{C}$ basis is mapped to the $\mathcal{B}$ basis using matrix P; then, the original representation A of T is applied to this vector; lastly, the result is mapped back into the $\mathcal{C}$ basis by P^{-1}.

5.3   Change of basis: linear map

Next, consider the general case of linear map $T : V \to W$ on vector spaces $(V^k, \mathcal{B})$ and $(W^m, \mathcal{C})$ where

$$\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$$

and

$$\mathcal{C} = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$$

. This linear map admits a matrix representation $A \in \mathbb{R}^{m \times k}$ such that

$$[T(\vec{x})]_{\mathcal{C}} = A[\vec{x}]_{\mathcal{B}}$$

This A has the form

$$A = \left[\begin{array}{cccc} | & | & & | \\ [T(\vec{v}_1)]_{\mathcal{C}} & [T(\vec{v}_2)]_{\mathcal{C}} & \cdots & [T(\vec{v}_k)]_{\mathcal{C}} \\ | & | & & | \end{array} \right]$$

Now consider changing bases of both vector spaces V and W: let $\mathcal{B}$ go to $\mathcal{B}'$ and $\mathcal{C}$ go to $\mathcal{C}'$ with change-of-basis matrices P and Q, respectively:

$$[\vec{x}]_{\mathcal{B}} = P[\vec{x}]_{\mathcal{B}'}, \ \vec{x} \in V$$

$$[\vec{y}]_{\mathcal{C}} = Q[\vec{y}]_{\mathcal{C}'}, \ \vec{y} \in W$$

Then using these relations, we see that we have

$$Q[T(\vec{x})]_{\mathcal{C}'} = AP[\vec{x}]_{\mathcal{B}'}$$

or

$$[T(\vec{x})]_{\mathcal{C}'} = Q^{-1}AP[\vec{x}]_{\mathcal{B}'}$$

so that the representation of our map T with respect to bases $\mathcal{B}$ and $\mathcal{C}$, A, becomes the matrix $Q^{-1}AP$ with respect to the bases $\mathcal{B}'$ and $\mathcal{C}'$.

In the special case where $T : V \to V$ is just a linear transformation on a space V with basis $\mathcal{B}$, then we can change to a new basis $\mathcal{B}'$ with the matrix P. If the map T has matrix representation A in the basis $\mathcal{B}$, then by the above discussion it has representation

$$[T(\vec{x})]_{\mathcal{B}'} = P^{-1}AP[\vec{x}]_{\mathcal{B}'}$$

in the $\mathcal{B}'$ basis. This agrees exactly with the case discussed in the previous section. It always behooves one to consider which basis and in which vector space every component of a matrix equation lies in order to verify that the function compositions are being performed correctly.

5.4   Change of basis: linear function composition

Chapter 6

Determinants

The determinant assigns to each square matrix a real number. If we can identify how the determinant behaves under elementary row operations and on the identity matrix, then we can understand how it operates on arbitrary nonsingular matrices, since these are reducible to the identity matrix via row operations.

(Lay out properties)

Theorem 6.1. *Let $A \in \mathbb{R}^{n \times n}$. Then A nonsingular if and only if $\det A \neq 0$.*

Proof. Assume A singular. Then in its row echelon form, U, contains a row of all zeroes, say row k. Then we have

$$\det (U) = 0 \cdot \det (U) = 0$$

Since A is obtained from U by a sequence of elementary row operations, each of which changes the determinant by a sign or scalar multiple, we must have $\det (A) = 0$, since neither a sign change or a scalar multiple will change 0. The contrapositive of this is that $\det A \neq 0 \implies A$ nonsingular.

Next, let A be nonsingular. Then it is row equivalent to the identity matrix, which has $\det I_n = 1$. Proceeding backward from the identity, we can reconstruct A by a sequence of row operations, each of which either changes the sign of the determinant or multiplies it by a nonzero scalar. Hence, we have $\det A \neq 0$. ∎

To actually perform practical computations, we can use the following:

Prop 6.2. *Let $A \in \mathbb{R}^{n \times n}$, and E be an elementary matrix. Then*

$$\det EA = \det A$$

Proof. We consider the possibilities for the elementary matrix E. First, if E encodes the scalar multiple of the kth row of A by arbitrary nonzero scalar c, then E has the

usual form

$$E = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & c & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix}$$

which is a scalar multiple of the identity matrix, so that $\det E = c \det I = c$. ∎

Theorem 6.3. *If $A, B \in \mathbb{R}^{n \times n}$, then*

$$\det AB = \det A \det B$$

Theorem 6.4. *If $A \in \mathbb{R}^{n \times n}$, then*

$$\det A^T = \det A$$

Chapter 7

Eigenvalues

Recall that the linear transformation $T : V \to V$ on a vector space V can be encoded in some basis $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ as a matrix such that

$$T(\vec{v}) = A[\vec{v}]_{\mathcal{B}}$$

That is, the action of T on V is a matrix-vector product. In another basis $\mathcal{B}'$, the matrix A will take on a different form, say A', but the function T itself is fundamentally unchanged; only its representation as a matrix has changed. Recall that a change of basis can be encoded in a matrix P so that

$$[\vec{v}]_{\mathcal{B}} = P[\vec{v}]_{\mathcal{B}'}$$

where P is the matrix consisting of columns which are the basis vectors in $\mathcal{B}'$ written in the basis $\mathcal{B}$. Given this change of basis of individual vectors, we have the relationship between the representations $[A]_{\mathcal{B}}$ and $[A]_{\mathcal{B}'}$ as

$$P[A]_{\mathcal{B}'} = [A]_{\mathcal{B}} P$$

where both sides operate on vectors in the $\mathcal{B}'$ basis, and return vectors in the $\mathcal{B}$ basis. Expressed differently, we have

$$[A]_{\mathcal{B}} = P[A]_{\mathcal{B}'} P^{-1}$$

A natural question to ask is if we can find a nice representation for the linear transformation $T : V \to V$ in some basis for $\mathcal{B}$. A particularly simple representation would be a diagonal one, where the matrix A is a diagonal matrix; in this case, T operates on $\vec{v} \in V$ by simply scaling each of its components:

$$A\vec{v} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_k \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \end{bmatrix} = \begin{bmatrix} \lambda_1 v_1 \\ \lambda_2 v_2 \\ \vdots \\ \lambda_k v_k \end{bmatrix}$$

Let $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a basis which admits this diagonal representation of T. In this

basis, the basis vectors are of the form

$$[\vec{v}_i]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

which is obvious by definition of these as basis vectors. Then note that

$$A\vec{v}_i = \lambda_i \vec{v}_i$$

This is the aforementioned component-wise scaling. This equation is always consistent, since the trivial solution always exists. We want to know if there are nonzero solutions to the linear system

$$(A - \lambda_i I)\,\vec{v}_i = 0$$

This solution set is

$$\mathrm{Nul}\,(A - \lambda_i I) = \{\vec{v} \in V \,|\, A\vec{v} = \lambda_i \vec{v}\}$$

In fact, we also want to identify the scalars λ_i which make this construction possible.

Recall that the null space of a matrix is nontrivial if and only if the matrix is singular, i.e., if the determinant of the matrix is zero. Thus, there will be some $\vec{v} \neq 0$ satisfying this equation if and only if

$$\det\,(A - \lambda_i I) = 0$$

This is the characteristic polynomial of the matrix A, denoted

$$p_A\,(t) = \det\,(A - tI)$$

and it is verified as a polynomial in t through minor-cofactor expansion.

Recall, though, that we are seeking a basis in which to express our linear transformation $T : V \to V$ as a diagonal matrix; this basis should be independent of any particular original basis, and so we need this polynomial to not depend on our choice of original basis. Since bases for a vector space are all related by similarity, the following suffices.

Prop 7.1. *Let A and B be similar matrices. Then $p_A\,(t) = p_B\,(t)$.*

Proof. Let $A = P^{-1}BP$. Then we have

$$\begin{aligned} p_A\,(t) &= \det\,(A - tI) \\ &= \det\,\left(P^{-1}BP - tI\right) \\ &= \det\,\left(P^{-1}\,(BP - tP)\right) \\ &= \det\,\left(P^{-1}\,(B - t)\,P\right) \\ &= \det\,(B - tI) \\ &= p_B\,(t) \end{aligned}$$

■

Thus, the characteristic polynomial for the transformation T is invariant under change of basis, and we can write it as $p_T(t)$. This agrees with what we expect.

We say that the roots of the characteristic polynomial are the *eigenvalues* of T, and the associated vectors are the eigenvectors. Notice that the eigenvector associated with any λ is not unique: in fact, the set

$$E(\lambda) = \text{Nul}(A - \lambda I)$$

can be shown to be a vector subspace spanned by $\vec{v}_i$. Further, if λ is an eigenvalue of matrix A, then there must be a nontrivial vector in $\text{Nul}(A - \lambda I)$, so eigenvectors always exist.

Prop 7.2. *Matrix A is singular if and only if A has 0 as an eigenvalue.*

Proof. If A has 0 as an eigenvalue, then

$$A\vec{v} = 0\vec{v} = 0$$

for some vector $\vec{v} \neq 0$. But then if A nonsingular, we have

$$\vec{v} = A^{-1}A\vec{v} = A^{-1}0 = 0$$

This is a contradiction, and so A must be singular.

Conversely, if A is nonsingular, then the equation

$$A\vec{v} = 0$$

has only the trivial solution, and so ■

Prop 7.3. *The eigenvalues of an upper or lower triangular matrix are its diagonal entries.*

Proof. This proof follows from the properties of determinants. Assume A is an upper triangular matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & a_{22} & & a_{2n} \\ & & \ddots & \vdots \\ & & & a_{nn} \end{bmatrix}$$

Then

$$A - \lambda I = \begin{bmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ & a_{22} - \lambda & & a_{2n} \\ & & \ddots & \vdots \\ & & & a_{nn} - \lambda \end{bmatrix}$$

Taking the determinant of this matrix, we obtain

$$\det(A - \lambda I) = (a_{11} - \lambda)\begin{vmatrix} a_{22} - \lambda & & a_{2n} \\ & \ddots & \vdots \\ & & a_{nn} - \lambda \end{vmatrix} = \cdots = \prod_{i=1}^{n}(a_{ii} - \lambda)$$

which is a polynomial in λ whose roots are exactly the diagonal entries. ∎

7.1 Distinct Eigenvalues

Theorem 7.4. *Let $T : V \to V$ be a linear transformation with eigenvectors $\vec{v}_1, \ldots, \vec{v}_k$ which has distinct eigenvalues $\lambda_1, \ldots, \lambda_k$. Then the eigenvectors are a linearly independent set.*

Proof. Assume by way of contradiction that the eigenvectors are not linearly dependent. Rearrange the eigenvectors so that the first m of them form a linearly independent set. Then $\vec{v}_1, \ldots, \vec{v}_m, \vec{v}_{m+1}$ are linearly dependent, so we have

$$\vec{v}_{m+1} = c_1 \vec{v}_1 + \cdots + c_m \vec{v}_m$$

for some choice of constants c_i. Then since these are eigenvectors of T, we have

$$
\begin{aligned}
0 &= (T - \lambda_{m+1} I) \vec{v}_{m+1} \\
&= (T - \lambda_{m+1} I)(c_1 \vec{v}_1 + \cdots + c_m \vec{v}_m) \\
&= (\lambda_1 c_1 \vec{v}_1 + \cdots + \lambda_m c_m \vec{v}_m) - \lambda_{m+1}(c_1 \vec{v}_1 + \cdots + c_m \vec{v}_m) \\
&= c_1 (\lambda_1 - \lambda_{m+1}) \vec{v}_1 + c_2 (\lambda_2 - \lambda_{m+1}) \vec{v}_2 + \ldots + c_m (\lambda_m - \lambda_{m+1}) \vec{v}_m
\end{aligned}
$$

By the linear independence of $\vec{v}_1, \ldots, \vec{v}_m$, we must have these coefficients

$$c_i (\lambda_i - \lambda_{m+1}) = 0$$

Further, since all eigenvalues are distinct, we need $c_i = 0$, and so $\vec{v}_{m+1} = 0$. But this is an eigenvector, and so we have reached a contradiction. Thus, The eigenvectors corresponding to distinct eigenvalues must be linearly independent. ∎

Corollary 7.4.1. *If V is an n-dimensional vector space and $T : V \to V$ has n distinct eigenvalues, then T is diagonalizable.*

Proof. If T has n distinct eigenvalues, then there are by the previous theorem n linearly independent eigenvectors of T. This is a maximal linearly independent subset of V and so it forms a basis for V. As we have shown previously, the representation of T in this eigenbasis is a diagonal matrix. ∎

When we have all distinct eigenvalues, the corresponding eigenspaces are one-dimensional. Note that the number of nonzero eigenvalues is equal to the rank of the matrix. This is because the matrix is similar to its diagonal representation in the eigenbasis, which has rank equal to the number of nonzero eigenvalues.

Chapter 8

Symmetric Matrices

Symmetric matrices have particularly nice eigenvector and -value properties. In particular, a symmetric matrix is guaranteed to have all real eigenvalues; it is also guaranteed to have an eigenbasis of orthonormal vectors, even if all eigenvalues are not distinct!

Theorem 8.1 (Spectral Theorem). *Let A be a symmetric $n \times n$ matrix. Then A has real eigenvalues and is diagonalizable by an orthogonal change-of-basis matrix:*

$$\Lambda = Q^{-1}AQ$$

where $Q^{-1} = Q^T$ and Λ is diagonal.

Given symmetric matrix A, we have the diagonalization

$$A = Q\Lambda Q^T = \begin{bmatrix} | & | & & | \\ \vec{q}_1 & \vec{q}_2 & & \vec{q}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} — & \vec{q}_1^T & — \\ — & \vec{q}_2^T & — \\ & \vdots & \\ — & \vec{q}_n^T & — \end{bmatrix}$$

$$= \begin{bmatrix} | & | & & | \\ \lambda_1\vec{q}_1 & \lambda_2\vec{q}_2 & & \lambda_n\vec{q}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} — & \vec{q}_1^T & — \\ — & \vec{q}_2^T & — \\ & \vdots & \\ — & \vec{q}_n^T & — \end{bmatrix}$$

What does a particular element of this product look like?

$$\begin{aligned} A_{ij} &= \sum_{k=1}^{n} \lambda_k \left(\vec{q}_k\right)_i \left(\vec{q}_k^T\right)_j \\ &= \lambda_1 \left(\vec{q}_1\right)_i \left(\vec{q}_1^T\right)_j + \lambda_2 \left(\vec{q}_2\right)_i \left(\vec{q}_2^T\right)_j + \cdots + \lambda_n \left(\vec{q}_n\right)_i \left(\vec{q}_n^T\right)_j \\ &= \lambda_1 \left(\vec{q}_1\vec{q}_1^T\right)_{ij} + \lambda_2 \left(\vec{q}_2\vec{q}_2^T\right)_{ij} + \cdots + \lambda_n \left(\vec{q}_n\vec{q}_n^T\right)_{ij} \\ &= \sum_{k=1}^{n} \lambda_k \left(\vec{q}_k\vec{q}_k^T\right)_{ij} \end{aligned}$$

so that

$$A = Q\Lambda Q^T = \sum_{k=1}^{n} \lambda_k \vec{q}_k \vec{q}_k^T$$

This is called the *spectral decomposition* of A. How does A act on a vector, $\vec{v}$, then? Note:

$$A\vec{v} = \sum_{k=1}^{n} \lambda_k \vec{q}_k \vec{q}_k^T \vec{x} = \sum_{k=1}^{n} \lambda_k \vec{q}_k \langle \vec{q}_k, \vec{v} \rangle$$

so that the action of A on a vector is the eigenvalue-weighted sum of the projection of $\vec{v}$ onto the eigenspace basis vectors (they are orthonormal).

(EXAMPLE) Consider the matrix

$$A = \begin{bmatrix} 6 & 2 \\ 2 & 9 \end{bmatrix}$$

It is real and symmetric, and so there is an orthogonal change-of-basis matrix Q such that the representation of A in this new basis is diagonal. Specifically, it is the eigen-basis. We can follow the usual prescription to obtain the eigenvalues and -vectors

$$\lambda_1 = 5, \ \vec{q}_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

$$\lambda_1 = 10, \ \vec{q}_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

These are easily orthonormal, and the change of basis matrix is

$$Q = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}$$

We can easily compute to see that

$$Q^T A Q = \begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix}$$

which is diagonal, as expected. We can also verify that the spectral decomposition returns the matrix A.

Note that as discussed in the chapter on inner products and orthogonality, the fact that the eigenbasis is orthonormal means that the usual dot product is preserved in this basis. To repeat the calculation, let $\vec{x}$ and $\vec{y}$ be orthogonal in the standard basis, where orthogonality is defined as

$$\langle \vec{x}, \vec{y} \rangle = \vec{x}^T \vec{y}$$

the usual dot product on $\mathbb{R}^2$. Then if we change the basis to the eigenbasis $\mathcal{Q}$, then we can write

$$[\vec{x}]_{\mathcal{Q}}^T [\vec{y}]_{\mathcal{Q}} = (Q\vec{x})^T (Q\vec{y}) = \vec{x}^T Q^T Q \vec{y} = \vec{x}^T \vec{y}$$

so that the dot product in the eigenbasis returns the same value as in the standard (original) basis. This is a consequence of the change-of-basis being an orthogonal transformation ($Q^{-1} = Q^T$), which itself results from the eigenvectors forming an orthogonal (even orthonormal) set in the usual dot product.

Chapter 9

Singular Values

Chapter 10

Inner Product Spaces

1. Definition of inner product. Extra structure on a vector space

2. Change of basis; orthogonality

3. Symmetric bilinear form. Matrix representation of inner product

10.1   Definition

An inner product on real vector space V (a vector space over the field $\mathbb{R}$) is a function

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$$

such that

1. $\langle \vec{u} + \vec{v}, \vec{w} \rangle = \langle \vec{u}, \vec{w} \rangle + \langle \vec{v}, \vec{w} \rangle$ for all $\vec{u}, \vec{v}, \vec{w} \in V$

2. $\langle \alpha \vec{u}, \vec{v} \rangle = \alpha \langle \vec{u}, \vec{v} \rangle$

3. $\langle \vec{u}, \vec{v} \rangle = \langle \vec{v}, \vec{u} \rangle$

4. $\langle \vec{u}, \vec{u} \rangle \geq 0$, and $\langle \vec{u}, \vec{u} \rangle = 0$ if and only if $\vec{u} = 0$

Together with this inner product, the vector space V is called an inner product space. The inner product gives rise to the familiar geometric notions of length and angle.

Note: an inner product space is automatically a metric space with metric

$$g(\vec{u}, \vec{v}) = \langle \vec{u} - \vec{v}, \vec{u} - \vec{v} \rangle$$

Note: an inner product space is automatically a normed space with norm

$$\|\vec{u}\| = \langle \vec{u}, \vec{u} \rangle^{1/2}$$

In the standard Euclidean basis $\mathcal{E}$ for $V \subset \mathbb{R}^n$, the standard Euclidean inner product is the familiar "dot product":

$$\langle \vec{x}, \vec{y} \rangle = [\vec{x}]_{\mathcal{E}}^T [\vec{y}]_{\mathcal{E}}$$

10.2 Norms, Schwartz Inequality

An inner product space is automatically a normed space; that is, given an inner product $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$, we have a natural norm

$$\|\cdot\| = \langle \cdot, \cdot \rangle^{1/2}$$

(DEFINE NORM)

The Cauchy-Schwarz inequality relates the magnitude of an inner product to the magnitude of the vectors:

Theorem 10.1 (Cauchy-Schwarz inequality)**.** *Let $\vec{v}$ and $\vec{w}$ be vectors in the inner product space V. Then*

$$|\langle \vec{v}, \vec{w} \rangle| \leq |\vec{v}|\,|\vec{w}|$$

Proof. In the case where one vector is 0, the equation holds trivially. If both are nonzero, then let $\lambda \in \mathbb{R}$ be arbitrary. Then consider the linear combination $\vec{v} + t\vec{w}$. We have

$$\begin{aligned}
0 &\leq \|\vec{v} + t\vec{w}\|^2 \\
&= \langle \vec{v} + t\vec{w}, \vec{v} + t\vec{w} \rangle \\
&= \langle \vec{v}, \vec{v} \rangle + 2t \langle \vec{v}, \vec{w} \rangle + t^2 \langle \vec{w}, \vec{w} \rangle \\
&= \|\vec{v}\|^2 + 2t \langle \vec{v}, \vec{w} \rangle + t^2 \|\vec{w}\|^2
\end{aligned}$$

which is the equation of a parabola in t. For this inequality to hold, we must have the discriminant in the quadratic formula

$$\frac{-2\langle \vec{v}, \vec{w} \rangle \pm \sqrt{4\langle \vec{v}, \vec{w} \rangle^2 - 4\|\vec{v}\|^2\|\vec{w}\|^2}}{2\|\vec{w}\|^2}$$

to be negative, i.e., there are no real roots to the parabola equation. This means

$$\langle \vec{v}, \vec{w} \rangle^2 - \|\vec{v}\|^2\|\vec{w}\|^2 \leq 0$$

or

$$|\langle \vec{v}, \vec{w} \rangle| \leq \|\vec{v}\|\,\|\vec{w}\|$$

$\blacksquare$

Given the Cauchy-Schwarz inequality (a consequence of the definition of inner product), we obtain the triangle inequality on the associated norm:

Theorem 10.2 (Triangle Inequality)**.** *Let $\vec{x}, \vec{y} \in V$ be vectors in an inner product space. Then*

$$\|\vec{x} + \vec{y}\| \le \|\vec{x}\| + \|\vec{y}\|$$

where $\|\cdot\|$ is the associated norm.

Thus, norms induced by inner products automatically satisfy the triangle inequality. Note that the triangle inequality is part of the definition of a norm, and so this proof demonstrates that the norm induced by an inner product is a bona fide norm.

Proof. The proof is simple, and follows from the Cauchy-Schwarz inequality:

$$\begin{aligned}
\|\vec{x} + \vec{y}\|^2 &= \langle \vec{x} + \vec{y}, \vec{x} + \vec{y} \rangle \\
&= \|\vec{x}\|^2 + 2\langle \vec{x}, \vec{y} \rangle + \|\vec{y}\|^2 \\
&\le \|\vec{x}\|^2 + 2|\langle \vec{x}, \vec{y} \rangle| + \|\vec{y}\|^2 \\
&\le \|\vec{x}\|^2 + 2\|\vec{x}\|\|\vec{y}\| + \|\vec{y}\|^2 \\
&= (\|\vec{x}\| + \|\vec{y}\|)^2
\end{aligned}$$

and so we have

$$\|\vec{x} + \vec{y}\| \le \|\vec{x}\| + \|\vec{y}\|$$

■

10.3 Change of Basis; Orthogonality

In general, the coordinate-based formula for an inner product of two vectors in a real vector space V is

$$\langle \vec{x}, \vec{y} \rangle = [\vec{x}]_{\mathcal{B}}^T Q [\vec{y}]_{\mathcal{B}}$$

where $\mathcal{B}$ is a basis for V and Q is a symmetric positive-definite matrix.

Let P be the change of basis from basis $\mathcal{E}$ to $\mathcal{C}$ of vector space $V \subset \mathbb{R}^n$, where $\mathcal{E}$ is the standard Euclidean basis. In the standard Euclidean basis $\mathcal{E}$, the usual Euclidean inner product has the form

$$\langle \vec{x}, \vec{y} \rangle = [\vec{x}]_{\mathcal{E}}^T [\vec{y}]_{\mathcal{E}}$$

This is the usual dot product of the coordinate vectors of $\vec{x}$ and $\vec{y}$. The change-of-basis matrix operates as

$$[\vec{x}]_{\mathcal{E}} = P[\vec{x}]_{\mathcal{C}}$$

Recall that this matrix P is built from columns which are the basis vectors of $\mathcal{C}$ expressed in terms of the basis vectors in $\mathcal{E}$. Then if we have an arbitrary inner product $\langle \cdot, \cdot \rangle$ on V in the basis $\mathcal{E}$,

$$\langle \vec{x}, \vec{y} \rangle = [\vec{x}]_{\mathcal{E}}^T Q [\vec{y}]_{\mathcal{E}}$$

then this expression changes as we move into the basis $\mathcal{C}$ as

$$\langle \vec{x}, \vec{y} \rangle = [\vec{x}]_{\mathcal{E}}^T Q [\vec{y}]_{\mathcal{E}} = \left(P^{-1}[\vec{x}]_{\mathcal{C}}\right)^T Q \left(P^{-1}[\vec{y}]_{\mathcal{C}}\right)$$

In particular, in the expression for the standard inner product, we have

$$\langle \vec{x}, \vec{y} \rangle = [\vec{x}]_{\mathcal{E}}^T [\vec{y}]_{\mathcal{E}} = \left(P [\vec{x}]_{\mathcal{C}} \right)^T \left(P [\vec{x}]_{\mathcal{C}} \right) = [\vec{x}]_{\mathcal{C}}^T P^T P [\vec{x}]_{\mathcal{C}}$$

which will be the standard Euclidean inner product in the $\mathcal{C}$ basis if and only if $P^T P = I$, i.e., if P is orthonormal. Since P has columns which are coordinates of $\mathcal{B}$ vectors in the standard basis $\mathcal{E}$. Hence, looking at what $P^T P$ computes, the standard Euclidean inner product of two vectors is equal to the dot product of the vector's coordinates if and only if the basis is orthonormal.

Intuition tells us that the orthogonality of vectors shouldn't change when their representation (basis) changes. However, on the contrary, the coordinate-based formula for the inner product — which defines orthogonality — is basis-dependent! This is seen by the previous paragraph.

The orthogonality of a basis (or any set of vectors) is dependent on an arbitrary inner product. A matrix A is said to be orthogonal if $A^T A = I$, i.e., it concerns the dot product of the columns of A. This is a bias toward orthonormality: let columns of A be coordinates of vectors in some orthonormal basis. In this case, $A^T A = I$ is equivalent to the column vectors being an orthonormal set. However, if we take an orthonormal set of vectors and assemble those vectors as columns of a matrix A, we don't necessarily have $A^T A = I$; this only happens if the coordinates are expressed relative to an orthonormal basis. On the other hand, we generally have some symmetric positive-definite matrix Q which defines our inner product. Then we always have $A^T Q A = 0$ when the columns of A are orthonormal with respect to the inner product.

The dot product on the space of coordinate vectors is not basis-independent, it is the unique inner product which makes the standard basis orthonormal. We can define a new inner product based on the dot product of the coordinate vectors of a given basis $\mathcal{B}$, and this will be the unique inner product that makes $\mathcal{B}$ orthonormal, but it will not agree with the original inner product unless $\mathcal{B}$ is already orthonormal, which we demonstrated above.

(EXAMPLE) Let

$$\mathcal{E} = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

and

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$$

be two basis for $\mathbb{R}^2$. Then the change of basis matrix P such that

$$[\vec{x}]_{\mathcal{E}} = P [\vec{x}]_{\mathcal{B}}$$

is

$$P = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

The columns of P are easily verified to be the vectors in $\mathcal{B}$ written in the basis $\mathcal{E}$. The standard inner product in the basis $\mathcal{E}$ (the standard basis) is

$$\langle \vec{x}, \vec{y} \rangle = [\vec{x}]_{\mathcal{E}}^T [\vec{y}]_{\mathcal{E}}$$

but we can rewrite this in terms of the $\mathcal{B}$-basis expansions of $\vec{x}$ and $\vec{y}$ as

$$\langle \vec{x}, \vec{y} \rangle = [\vec{x}]_{\mathcal{B}}^T P^T P [\vec{y}]_{\mathcal{B}}$$

which is easily verified on some toy examples. Only with this modification to the inner product does the change of basis from $\mathcal{E}$ to $\mathcal{B}$ preserve orthogonality in the sense of the usual inner product on $\mathcal{E}$ (the "dot product").

Chapter 11

Jordan Canonical Form

Chapter 12

Dual Spaces

1. Definition

2.