Chapter 4 & 5: Vector Spaces & Linear Transformations

Philip Gressman

University of Pennsylvania
The purpose of Chapter 4 is to think **abstractly** about vectors.

- We will identify certain fundamental laws which are true for the kind of vectors (i.e., column vectors) that we already know.
- We will consider the abstract notion of a **vector space**.
- The purpose of doing this is to apply the theory that we already know and love to situations which appear **superficially different** but are really very similar from the right perspective.
- The **main application** for us will be solutions of linear ordinary differential equations. Even though they are functions instead of column vectors, they behave in exactly the same way and can be understood completely.
Abstract Setup

We begin with a collection $V$ of things we call vectors and a collection of scalars $F$ (which will be either real numbers, rational numbers, or complex numbers usually).

A1 We are given a formula for vector addition $u + v$. The sum of two things in $V$ will be another thing in $V$.

A2 We are given a formula for scalar multiplication $cv$ for $c \in F$ and $v \in V$. It should be another thing in $V$.

A3 (Commutativity) It should be true that $u + v = v + u$ for any $u, v \in V$.

A4 (Associativity of +) It should be true that $(u + v) + w = u + (v + w)$ for any $u, v, w \in V$.

A5 (Additive identity) There should be a vector called 0 such that $v + 0 = v$ for any $v \in V$.

A6 (Additive inverses) For any $v \in V$, there should be a vector $-v$ so that $v + (-v) = 0$.

A7 (Multiplicative identity) For any $v \in V$, we should have $1v = v$. 

Philip Gressman
Abstract Setup (continued)

A8  (Associativity of \( \cdot \)) For any \( v \in V \) and \( c_1, c_2 \in F \), we should have \( (c_1 c_2)v = c_1(c_2 v) \).

A9  (Distributivity I) For any \( u, v \in V \) and \( c \in F \), it should be true that
\[
c(u + v) = cu + cv.
\]

A10 (Distributivity II) For any \( c_1, c_2 \in F \) and \( u \in V \), it should be true that
\[
(c_1 + c_2)u = c_1 u + c_2 u.
\]

The Point: Someone hands you a set of things \( V \) they call vectors and a set \( F \) of scalars. If they call \( V \) a vector space over \( F \), they mean that they have checked that A1 through A10 are always true.

If someone asks you if a given \( V \) is a vector space over \( F \), you must check that each one of A1 through A10 is true under all circumstances.
Vector Spaces: Examples

- Real-valued convergent sequences \( \{a_n\} \) are a vector space over \( \mathbb{R} \) when \( \{a_n\} + \{b_n\} \) is the sequence \( c_n := a_n + b_n \) and scalar multiplication is done termwise. They are not a vector space over \( \mathbb{C} \). Complex-valued convergent sequences \( \{b_n\} \) are a vector space over \( \mathbb{C} \). They are also a vector space over \( \mathbb{R} \).

- Real-valued functions \( f \) on an open interval \( I \) such that \( f, f', f'', \ldots, f^{(k)} \) are all continuous functions on \( I \) form a vector space under pointwise addition and pointwise scalar multiplication. It’s called \( C^k(I) \).

- Item polynomials with real coefficients are a vector space over \( \mathbb{R} \) under the usual notions of addition and scalar multiplication.
Any two things which are vector spaces over $F$ will have many things in common, far beyond just the properties A1 through A10. For example:

- $0u = 0$ for any $u \in V$. Why? Well, $0u + u = (0 + 1)u = 1u = u$, but now add $-u$ to both sides.

- There is only one vector 0 which always works for A5. That’s because $v = v + 0’ = 0’ + v$; now add 0 to each side.

- $c0 = 0$ for any $c \in F$. Why? If $c \neq 1$, notice that $\frac{1}{c-1}0 + 0 = \frac{1}{c-1}0$. Un-distribute the left-hand side and then multiply both sides by $(c-1)$. If $c = 1$ then axiom A7 already works.

See Theorem 4.2.6 for several more examples.
A **vector subspace** $S$ is a subset of $V$ which is a vector space in its own right. The conditions to check are A1 and A2 (namely, that sums of things in $S$ need to remain in $S$ and scalar multiples of things in $S$ still belong in $S$). All the other conditions will still be true because they were true in $V$.

**Important Example**

If you have **any** homogeneous system of equations in $n$ unknowns, then the solution set of the system will be a subspace of $F^n$ (that’s the notation for $n$-tuples of things in $F$).

**Second Important Example**

Solutions of any linear homogeneous ODE form a vector subspace of the vector space of all functions on the real line (or on a finite interval $I$ if you wish).
For systems of equations with \( m \) unknowns, solutions of any homogeneous system of equations always form a vector subspace of \( \mathbb{R}^m \) (or \( \mathbb{C}^m \) if solutions are allowed to be complex). Given an \( n \times m \) matrix \( A \), we define the \textbf{null space} or \textbf{kernel} of \( A \) to be those vectors \( x \) in \( \mathbb{R}^m \) (or \( \mathbb{C}^m \)) such that \( Ax = 0 \).
Linear Combinations

If you are given vectors $v_1, v_2, \ldots, v_k$, any vector $w$ which can be written in the form

$$c_1 v_1 + c_2 v_2 + \cdots + c_k v_k$$

for scalars $c_1, c_2, \ldots, c_k$ is said to be a **linear combination** of $v_1, \ldots, v_k$.

**Definition**

Given any list of vectors $v_1, \ldots, v_k \in V$, the set of all linear combinations of $v_1, \ldots, v_k$ is a vector subspace of $V$. It is called the **span** of $v_1, \ldots, v_k$. 
Types of Problems

• Determine whether a given set of vectors spans a vector space.
  • A collection of fewer than $n$ vectors in $\mathbb{R}^n$ never spans all of $\mathbb{R}^n$.
  • A collection of exactly $n$ vectors in $\mathbb{R}^n$ spans all of $\mathbb{R}^n$ if and only if the determinant of the matrix they make by adjoining columns is nonzero.
  • A collection of more than $n$ row vectors spans $\mathbb{R}^n$ if and only if the rank of the matrix made by adjoining rows is $n$.

• Given a list of vectors, classify all vectors in the span.
  • Write a system of equations with unknowns equal to the coefficients $c_j$ in the linear combination. Try to solve the system for an arbitrary vector $b$ solving $b = c_1 v_1 + \cdots + c_k v_k$.

• Determine whether the spans of two sets are the same or different.

• Identify whether a particular vector is in the span or not.
Identify when a specific vector $y$ is in the span of $v_1, \ldots, v_k$? Determine if these equations can be solved:

$$
\begin{bmatrix}
  v_1 & \cdots & v_k
\end{bmatrix}
\begin{bmatrix}
  c_1 \\
  \vdots \\
  c_k
\end{bmatrix} = 
\begin{bmatrix}
  y
\end{bmatrix}
$$

Classify the span of $v_1, \ldots, v_k$? Treat the entries of $y$ as variables, determine what constraints (if any) on the entries of $y$ are needed to make the equations consistent.
Two Ways to Describe Subspaces

There are two basic ways to give complete descriptions of a vector subspace:

- Give a list of vectors that span the vector subspace; in other words, find a collection of vectors which is (1) small enough to belong to the subspace, and (2) big enough so that they don’t belong to a smaller subspace.

- Give a complete list of linear constraints (i.e., equations) that all vectors in the subspace must satisfy.

In both cases, it is important to know what the minimal number of objects (either vectors or constraints) must be—there’s not much point in working harder than is necessary.
Linear Dependence

A collection of vectors $v_1, \ldots, v_k$ is said to be **linearly dependent** when it’s possible to write 0 as a linear combination of $v_1, \ldots, v_k$ which is nontrivial, i.e.,

$$c_1 v_1 + \cdots + c_k v_k = 0$$

and not every one of $c_1, \ldots, c_k$ is zero.

Linear dependence of a set of vectors means that one or more vectors may be removed without changing their span. Linear dependence means that the system

$$
\begin{bmatrix}
  v_1 & \cdots & v_k
\end{bmatrix}
\begin{bmatrix}
  c_1 \\
  \vdots \\
  c_k
\end{bmatrix}
= 
\begin{bmatrix}
  0 \\
  \vdots \\
  0
\end{bmatrix}
$$

has nontrivial solutions.
Definition

A *basis* of a vector space $V$ is a collection of vectors $v_1, \ldots, v_k$ which is linearly independent and spans $V$. **In other words,**

$$c_1 v_1 + c_2 v_2 + \cdots + c_k v_k = y$$

has a unique solution for any choice of $y \in V$ (i.e., always one solution, never more than one).

The column vector

$$\begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix}$$

is called the *coordinates* of $y$ with respect to the *ordered basis* $v_1, \ldots, v_k$. 
Bases

• The **standard basis** of $\mathbb{R}^n$ is given by

\[
\begin{bmatrix}
1 \\
0 \\
\vdots \\
0
\end{bmatrix},
\begin{bmatrix}
0 \\
1 \\
\vdots \\
0
\end{bmatrix}, \ldots,
\begin{bmatrix}
0 \\
0 \\
\vdots \\
1
\end{bmatrix}
\]

• Bases of the same vector space must always have the same number of vectors. This is called the **dimension** of the vector space. **ANY** basis of $\mathbb{R}^n$ must have $n$ vectors.

• If the dim. = $n$, then any list of more than $n$ vectors must be linearly dependent.

• If dim. = $n$, then any list of fewer than $n$ vectors never spans.

• If you **know** you have the right **number** of vectors, then linear independence and spanning go hand-in-hand (i.e., if one is true then the other is true; if one is false then the other is false).
• Coordinates change when the basis changes; the effect of changing bases is encoded in the change of basis matrix. The columns of the change of basis matrix are the coordinates in the new basis of the old basis vectors.

• Basic Skills:
  • How do you verify that a given set is a basis of a vector space?
  • How do you discover a basis for a given vector space?
  • How do you convert coordinates from one basis to another?
What are the mechanical steps to check for linear independence of a list of vectors $v_1, \ldots, v_k$:

- If they all belong to some vector space $V$ of dimension $< k$, then stop: they cannot be linearly independent.
- If they belong to $\mathbb{R}^n$ or $\mathbb{C}^n$, treat the statement
  \[ c_1 v_1 + \cdots + c_k v_k = 0 \]
as an equation to solve for unknowns $c_1, \ldots, c_k$. Solve by row reduction like normal; if there is only one solution then they are linearly independent.
- If they belong to an exotic vector space but you happen to know a basis, write everything out in coordinates. This reduces the computational problem to the previous case.
- If you do not know a nice basis, e.g. $C^\infty(I)$, you have to be more clever. In this case, one option is to use what is known as the Wronskian.
**Theorem**

For any $m \times n$ matrix $A$,

$$\text{rank}(A) + \text{nullity}(A) = n.$$ 

In other words, the rank of $A$ plus the dimension of the null space of $A$ equals the number of columns.

Remember:

- The rank of the matrix equals the number of nonzero rows after row reduction.
- The rank is also the dimension of the row space.
- The rank is also the dimension of the span of the columns.
Applications of Rank-Nullity

Analysis of the equations $Ax = 0$ when $A$ is an $m \times n$ matrix:

- If the rank of $A$ is $n$, then the only solution is $x = 0$.
- Otherwise if $r$ is the rank, then the nullity is $n - r$, meaning that all solutions have the form

$$x = c_1x_1 + \cdots + c_{n-r}x_{n-r}$$

for any linearly-independent set of solutions $x_1, \ldots, x_{n-r}$.

Analysis of $Ax = b$:

- If $b$ is not in the column space, then there is no solution.
- If $b$ is in the column space and $\text{rank}(A) = n$, then there is a unique solution.
- If $b$ is in the column space and $\text{rank}(A) = r < n$, then

$$x = c_1x_1 + \cdots + c_{n-r}x_{n-r} + x_p$$

where $x_1, \ldots, x_{n-r}$ are linearly-independent solutions of $Ax = 0$ and $x_p$ is any single solution of $Ax = b$. 
A **linear transformation** is any mapping $T : V \to W$ between vector spaces $V$ and $W$ over the same field which satisfies

$$T(u + v) = Tu + Tv \quad \text{and} \quad T(cv) = c(Tv)$$

for any vectors $u, v \in V$ and any scalar $c$.

The canonical example is when $T$ is multiplication by some given matrix $A$, but there are a variety of things which are linear transformations.
Suppose $T$ is a linear transformation of a vector space $V$ with ordered basis $B$ into a vector space $W$ with ordered basis $C$.

**Basic Question**

If $v \in V$ has coordinates

$$
\begin{bmatrix}
  c_1 \\
  \vdots \\
  c_n
\end{bmatrix}
$$

$B$

what will be the coordinates

$$
\begin{bmatrix}
  d_1 \\
  \vdots \\
  d_m
\end{bmatrix}
$$

$C$

of $Tv$?
Geometry of Matrix Multiplication

To understand the importance of eigenvectors and eigenvalues (§5.6–8), it is first necessary to understand the geometric meaning of matrix multiplication. If you visit the link http://www.math.hawaii.edu/phototranspage.html, you can explore the result of applying any $2 \times 2$ matrix you choose to any image of your choice.

Geometry of Matrix Multiplication

\[
\begin{bmatrix}
  x' \\
  y'
\end{bmatrix} =
\begin{bmatrix}
  0.5 & 0 \\
  0 & 1
\end{bmatrix}
\begin{bmatrix}
  x \\
  y
\end{bmatrix}
\]

Left sheep in $xy$-plane  
Right sheep in $x'y'$-plane
Geometry of Matrix Multiplication

\[
\begin{bmatrix}
  x' \\
  y'
\end{bmatrix} = \begin{bmatrix}
  1 & 0 \\
  0 & 1.8
\end{bmatrix} \begin{bmatrix}
  x \\
  y
\end{bmatrix}
\]

Left sheep in \( xy \)-plane

Right sheep in \( x'y' \)-plane
Geometry of Matrix Multiplication

\[
\begin{bmatrix}
    x' \\
    y'
\end{bmatrix} = \begin{bmatrix}
    1 & 0.7 \\
    0 & 1
\end{bmatrix} \begin{bmatrix}
    x \\
    y
\end{bmatrix}
\]

Left sheep in \(xy\)-plane

Right sheep in \(x'y'\)-plane
Geometry of Matrix Multiplication

\[
\begin{bmatrix}
x' \\
y'
\end{bmatrix} = \begin{bmatrix}
1 & 0 \\
-0.6 & 1
\end{bmatrix} \begin{bmatrix}
x \\
y
\end{bmatrix}
\]

Left sheep in \(xy\)-plane

Right sheep in \(x'y'\)-plane
Geometry of Matrix Multiplication

\[
\begin{bmatrix}
  x' \\
  y'
\end{bmatrix} = \begin{bmatrix}
  1.2 & 0.5 \\
  0.5 & 1.2
\end{bmatrix} \begin{bmatrix}
  x \\
  y
\end{bmatrix}
\]

Left sheep in \( xy \)-plane

Right sheep in \( x'y' \)-plane
Eigenvectors

**Eigenvectors and Eigenvalues**

For a square matrix $A$, if there is a (nonzero) column vector $E$ and a number $\lambda$ so that

$$AE = \lambda E$$

then we call $E$ an eigenvector and $\lambda$ its eigenvalue.

Geometrically, the existence of an eigenvector means that there is a direction which is preserved by the matrix $A$: in the same direction as $E$, the only effect of multiplication by $A$ is to stretch (or shrink, or reflect) by a factor $\lambda$.

The nicest possible situation is when an $n \times n$ matrix has $n$ linearly independent eigenvectors. This would mean that the matrix $A$ can be understood solely in terms of stretchings in each of the various directions.

In practice, this does not always happen.
Finding the Eigenvalues

Characteristic Equation

If $A$ is an $n \times n$ matrix, then the determinant

$$\det(A - \lambda I)$$

is a polynomial of degree $n$ in terms of $\lambda$ (called the characteristic polynomial). The eigenvalues of $A$ must be solutions of the equation

$$\det(A - \lambda I) = 0.$$ 

Why? Because $\det(A - \lambda I) = 0$ if and only if there is a nonzero column vector $E$ for which $(A - \lambda I)E = 0$. This means $E$ is an eigenvector with eigenvalue $A$.

Fact

For triangular matrices, the eigenvalues are just the diagonal entries.
Finding the Eigenvectors

Solving for $E$

Once you know the roots of the characteristic polynomial, the way that you find eigenvalues is to solve the system of equations

$$(A - \lambda_j I)X = 0$$

where you fix $\lambda_j$ to be any root. Any solution $X$ is an eigenvector.

Fact

Assuming you’ve solved for the eigenvalues correctly, there will always be at least one solution of this system (i.e., at least one eigenvector). Sometimes there will be multiple eigenvectors. You should always choose them to be linearly independent (and remember that the number of identically zero rows you find when row reducing $A - \lambda I$ will tell you exactly how many eigenvectors to find).
Points to Remember

**Not Enough Eigenvectors**

Sometimes an $n \times n$ matrix may not have $n$ linearly independent eigenvectors (eigenvectors with different eigenvalues are automatically linearly independent). This only happens when the characteristic polynomial has repeated roots, and even then it only happens some of the time.

**Complex Eigenvalues**

If there are complex roots to the characteristic equation, you will get complex eigenvalues. If your matrix had all real entries, then complex eigenvalues will always appear in conjugate pairs. Moreover, if you find an eigenvector for one of these eigenvalues, you can automatically construct an eigenvector for the conjugate eigenvalue by taking the complex conjugate of the eigenvector itself.
Suppose $V_1, \ldots, V_n$ are column vectors. Linear combinations of these vectors are vectors which can be written in terms of the formula

$$W = c_1 V_1 + \cdots + c_n V_n$$

for some scalars $c_1, \ldots, c_n$. We can write the vectors as column vectors; the formula becomes

$$
\begin{bmatrix}
| & | & & | \\
V_1 & \cdots & | & V_n \\
| & | & & | \\
c_1 & \vdots & & c_n \\
| & | & & | \\
\end{bmatrix}
= 
\begin{bmatrix}
| \\
W \\
|
\end{bmatrix}.
$$

If $V_1, \ldots, V_n$ are linearly independent vectors in $n$-dimensional space, then the matrix formed by concatenating the column vectors (call it $P$) will be invertible.
Diagonalization

In short, if \( W \) is a column vector and \( P \) is a matrix whose columns are linearly independent vectors \( V_1, \ldots, V_n \), then the column vector 

\[
P^{-1}W
\]

tells you the unique choice of coefficients \( c_1, \ldots, c_n \) necessary to make \( W = c_1 V_1 + \cdots + c_n V_n \).

Next, suppose that \( V_1, \ldots, V_n \) happen to be eigenvectors of a matrix \( A \) (with eigenvalues \( \lambda_1, \ldots, \lambda_n \)).

\[
W = c_1 V_1 + \cdots + c_n V_n \Rightarrow AW = c_1 \lambda_1 V_1 + \cdots + c_n \lambda_n V_n.
\]

In other words,

\[
P^{-1}AW = \begin{bmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & \cdots & \vdots \\
\vdots & \cdots & \cdots & 0 \\
0 & \cdots & 0 & \lambda_n
\end{bmatrix}
\]

\[
P^{-1}W
\]
Diagonalization

We conclude: If $A$ is an $n \times n$ matrix which happens to have $n$ linearly independent eigenvectors, let $P$ be the matrix whose columns are those eigenvectors and let $D$ be a diagonal matrix made of the eigenvalues of $A$. We conclude $P^{-1}AW = DP^{-1}W$ for any column vector $W$ you like. Consequently

$$A = PDP^{-1} \text{ and } D = P^{-1}AP.$$ 

We call $D$ the **diagonalization** of $A$. Having a diagonalization $A$ means that in some appropriate basis, the matrix $A$ just acts like a coordinate-wise scaling. 

**A sufficient condition** to be diagonalizable is when $A$ has $n$ distinct eigenvalues. **High powers of** $A$ are easy to compute with diagonalizations:

$$A^N = PD^NP^{-1}$$
• If $A$ is symmetric, then it may always be diagonalized, and it will always be true in this case that the eigenvectors can be chosen so that they’re mutually orthogonal.

• Diagonalizations are useful to reduce the complexity of problems (since operating with diagonal matrices is generally much easier than it would be with more general matrices).

• Applications include classification of conic sections and solutions of systems of linear ordinary differential equations.
Diagonalization: Further Examples

- Classification and Transformation of Conic Sections
- Logistics / Planning / Random Walks
- Computing Functions of a Matrix
- Data Analysis: http://en.wikipedia.org/wiki/Principal_component_analysis
- Quantum Mechanics...
Some matrices / linear transformations cannot be diagonalized. Examples:

\[
\begin{bmatrix}
1 & 2 & 3 \\
0 & 1 & 2 \\
0 & 0 & 1
\end{bmatrix}
\]
or Differentiation on \( P_n, n \geq 1 \)

In this case we can do a next best option and find what is called the **Jordan Canonical Form**.

A **Jordan Block** corresponding to \( \lambda \) is a square matrix with \( \lambda \) in each diagonal entry, 1 in every entry immediately above the diagonal, and zero elsewhere.
For any fixed eigenvalue $\lambda_0$:

- The number of blocks equals the number of eigenvectors, which equals the nullity of $A - \lambda_0 I$.
- The sum of the sizes of all blocks associated to $\lambda_0$ equals the power of $(\lambda - \lambda_0)$ found in the characteristic polynomial.

Example: If $\det(A - \lambda I) = (\lambda - 1)(\lambda - 2)^3$ then we can say that

- $A$ is a $4 \times 4$ matrix
- There is one eigenvector with eigenvalue 1.
- As for eigenvalue 2, there are either
  - Three eigenvectors
  - Two eigenvectors: $1 \times 1$ and $2 \times 2$ Jordan blocks
  - One eigenvector: $3 \times 3$ Jordan block
Some Examples to Think About

Example 1: \( \det(A - \lambda I) = (\lambda - 1)(\lambda - 2)^2(\lambda - 3)^3 \), nullity of \( A - 2I \) is 2 and nullity of \( A - 3I \) is also 2.

Example 2: \( \det(A - \lambda I) = (\lambda - 1)^2(\lambda - 2)^2(\lambda - 3)^4 \), nullity of \( A - I \), \( A - 2I \), and \( A - 3I \) are all 2.
Generalized Eigenvectors

**Definition**

A vector \( v \neq 0 \) is called a **generalized eigenvector** with eigenvalue \( \lambda \) when \((A - \lambda I)^p v = 0\) for some \( p \).

If \((A - \lambda I)^p v = 0\), then automatically \((A - \lambda I)^{p+1} v = 0\) and so on for all higher powers. Thus if \( v \) is specified, we generally need to know what the smallest value of \( p \) is that makes the formula true.

To build a basis in which the matrix takes its Jordan Form, we start with vectors \( v \) such that \((A - \lambda I)^p v = 0 \) and \( p \) is minimal; then we put the vectors

\[(A - \lambda I)^{p-1} v, (A - \lambda I)^{p-2} v, \ldots, (A - \lambda I) v, v\]

in *that order* into the basis.