STEPHEN C. NEWMAN

SEMI-RIEMANNIAN GEOMETRY

THE MATHEMATICAL LANGUAGE OF GENERAL RELATIVITY

WILEY

Semi-Riemannian Geometry

Semi-Riemannian Geometry

The Mathematical Language of General Relativity

STEPHEN C. NEWMAN
University of Alberta
Edmonton, Alberta, Canada

WILEY

This edition first published 2019

Registered Office
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

Editorial Office
111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data

Names: Newman, Stephen C., 1952- author.

Title: Semi-Riemannian geometry : the mathematical language of general relativity / Stephen C. Newman (University of Alberta, Edmonton, Alberta, Canada).

Description: Hoboken, New Jersey : Wiley, [2019] | Includes bibliographical references and index. |

Identifiers: LCCN 2019011644 (print) | LCCN 2019016822 (ebook) | ISBN 9781119517542 (Adobe PDF) | ISBN 9781119517559 (ePub) | ISBN 9781119517535 (hardcover)

Subjects: LCSH: Semi-Riemannian geometry. | Geometry, Riemannian. | Manifolds (Mathematics) | Geometry, Differential.

Classification: LCC QA671 (ebook) | LCC QA671 .N49 2019 (print) | DDC 516.3/73–dc23

LC record available at https://lccn.loc.gov/2019011644

Cover design: Wiley

Set in 10/12pt Computer Modern by SPi Global, Chennai, India

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To Sandra

Contents

Preface

Physics texts on general relativity usually devote several chapters to an overview of semi-Riemannian geometry. Of necessity, the treatment is cursory, covering only the essential elements and typically omitting proofs of theorems. For physics students wanting greater mathematical rigor, there are surprisingly few options. Modern mathematical treatments of semi-Riemannian geometry require grounding in the theory of curves and surfaces, smooth manifolds, and Riemannian geometry. There are numerous books on these topics, several of which are included in Further Reading. Some of them provide a limited amount of material on semi-Riemannian geometry, but there is really only one mathematics text currently available that is devoted to semi-Riemannian geometry and geared toward general relativity, namely, *Semi-Riemannian Geometry: With Applications to Relativity* by Barrett O'Neill. This is a classic, but it is pitched at an advanced level, making it of limited value to the beginner. I wrote the present book with the aim of filling this void in the literature.

There are three parts to the book. Part I and the Appendices present background material on linear algebra, multilinear algebra, abstract algebra, topology, and real analysis. The aim is to make the book as self-contained as possible. Part II discusses aspects of the classical theory of curves and surfaces, but differs from most other expositions in that Lorentz as well as Euclidean signatures are discussed. Part III covers the basics of smooth manifolds, smooth manifolds with boundary, smooth manifolds with a connection, and semi-Riemannian manifolds. It concludes with applications to Lorentz vector spaces, Maxwell's equations, and the Einstein tensor. Not all theorems are provided with a proof, otherwise an already lengthy volume would be even longer.

The manuscript was typed using the WYSIWYG scientific word processor EXP®, and formatted as a camera-ready PDF file using the open-source TEX-LATEX typesetting system MiKTeX, available at `https://miktex.org`. Figure 19.5.1 was prepared using the TEX macro package `diagrams.sty` developed by Paul Taylor. I am indebted to Professor John Lee of the University of Washington for reviewing portions of the manuscript. Any remaining errors or deficiencies are, of course, solely my responsibility.

I am most interested in receiving your comments, which can be emailed to me at `stephen.newman@ualberta.ca`. A list of corrections will be posted on the website `https://sites.ualberta.ca/~sn2/`. Should the email address become unavailable, an alternative will be included with the list of corrections.

On the other hand, if the website becomes inaccessible, the list of corrections will be stored as a public file on Google Drive that can be searched using “Corrections to Semi-Riemannian Geometry by Stephen Newman”.

Allow me to close by thanking my wife, Sandra, for her unwavering support and encouragement throughout the writing of the manuscript. It is to her, with love, that this book is dedicated.

Part I

Preliminaries

Differential geometry rests on the twin pillars of linear algebra–multilinear algebra and topology–analysis. Part I of the book provides an overview of selected topics from these areas of mathematics. Most of the linear algebra presented here is likely familiar to the reader, but the same may not be true of the multilinear algebra, with the exception of the material on determinants. Topology and analysis are vast subjects, and only the barest of essentials are touched on here. In order to keep the book to a manageable size, not all theorems are provided with a proof, a remark that also applies to Part II and Part III.

Chapter 1

Vector Spaces

1.1 Vector Spaces

The definition of a **vector space** over a field and that of a **subspace** of a vector space are given in Section B.6. Our focus in this book is exclusively on vector spaces over the real numbers (as opposed to the complex numbers or some other field).

Throughout, all vector spaces are over $\mathbb{R}$, the field of real numbers.

For brevity, we will drop the reference to $\mathbb{R}$ whenever possible and write, for example, “linear” instead of “$\mathbb{R}$-linear”.

Of particular importance is the vector space $\mathbb{R}^m$, but many other examples of vector spaces will be encountered. It is easily shown that the intersection of any collection of subspaces of a vector space is itself a subspace. The **zero vector** of a vector space is denoted by 0, and the **zero subspace** of a vector space by $\{0\}$. The **zero vector space**, also denoted by $\{0\}$, is the vector space consisting only of the zero vector. We will generally avoid explicit consideration of the zero vector space. Most of the results on vector spaces either apply directly to the zero vector space or can be made applicable with a minor reworking of definitions and proofs. The details are usually left to the reader.

Example 1.1.1. Let V and W be vector spaces. Following Section B.5 and Section B.6, we denote by $\mathrm{Lin}(V, W)$ the vector space of linear maps from V to W, where addition and scalar multiplication are defined as follows: for all maps A, B in $\mathrm{Lin}(V, W)$ and all real numbers c,

$$(A + B)(v) = A(v) + B(v)$$

and

$$(cA)(v) = cA(v)$$

Semi-Riemannian Geometry, First Edition. Stephen C. Newman.

for all vectors v in V. The zero element of $\mathrm{Lin}(V, W)$, denoted by 0, is the zero map, that is, the map that sends all vectors in V to the zero vector 0 in W. When $V = W$, we make $\mathrm{Lin}(V, V)$ into a ring by defining multiplication to be composition of maps: for all maps A, B in $\mathrm{Lin}(V, V)$, let

$$A \circ B(v) = A\big(B(v)\big)$$

for all vectors v in V. The identity element of the ring $\mathrm{Lin}(V, V)$ is the identity map on V, denoted by id_V. $\Diamond$

A **linear combination** of vectors in a vector space V is defined to be a *finite* sum of the form $a^1 v_1 + \cdots + a^k v_k$, where $a^1, \ldots, a^k$ are real numbers and $v_1, \ldots, v_k$ are vectors in V. The possibility that some (or all) of $a^1, \ldots, a^k$ equal zero is not excluded.

Let us pause here to comment on an aspect of notation. Following the usual convention in differential geometry, we index the scalars and vectors in a linear combination with superscripts and subscripts, respectively. This opens the door to the **Einstein summation convention**, according to which, for example, $a^1 v_1 + \cdots + a^k v_k$ and $\sum_{i=1}^{k} a^i v_i$ are abbreviated as $a^i v_i$. The logic is that when an expression has a superscript and subscript in common, it is understood that the index is being summed over. Despite the potential advantages of this notation, especially when multiple indices involved, the Einstein summation convention will *not* be adopted here.

Let S be a (nonempty and not necessarily finite) subset of V. The **span of S** is denoted by $\mathrm{span}(S)$ and defined to be the set of linear combinations of vectors in S:

$$\mathrm{span}(S) = \{a^1 v_1 + \cdots + a^k v_k : a^1, \ldots, a^k \in \mathbb{R}; \\ v_1, \ldots, v_k \in S; k = 1, 2, \ldots\}.$$

For a vector v in V, let us denote

$$\mathrm{span}(\{v\}) = \{av : a \in \mathbb{R}\} \qquad \text{by} \qquad \mathbb{R}v.$$

For example, in $\mathbb{R}^2$, we have

$$\mathrm{span}\big(\{(1, 0), (0, 1)\}\big) = \mathbb{R}^2$$

and

$$\mathrm{span}\big(\{(1, 0)\}\big) = \mathbb{R}(1, 0) = \{(a, 0) \in \mathbb{R}^2 : a \in \mathbb{R}\}.$$

It is easily shown that $\mathrm{span}(S)$ is a subspace of V. In fact, $\mathrm{span}(S)$ is the smallest subspace of V containing S, in the sense that any subspace of V containing S also contains $\mathrm{span}(S)$. When $\mathrm{span}(S) = V$, it is said that **S spans V** or that the **vectors in S span V**, and that each vector in V is in the **span of S**.

We say that **S is linearly independent** or that the **vectors in S are linearly independent** if the only linear combination of *distinct* vectors in S

that equals the zero vector is the one with all coefficients equal to 0. That is, if $v_1, \dots, v_k$ are distinct vectors in S and $a^1, \dots, a^k$ are real numbers such that $a^1 v_1 + \cdots + a^k v_k = 0$, then $a^1 = \cdots = a^k = 0$. Evidently, any subset of a linearly independent set is linearly independent. When S is not linearly independent, it is said to be **linearly dependent**. In particular, the zero vector in any vector space is linearly dependent. As further examples, the vectors $(1,0),(0,1)$ in $\mathbb{R}^2$ are linearly independent, whereas $(0,0),(1,0)$ and $(1,0),(2,0)$ are linearly dependent.

The next result shows that when a linearly independent set does not span a vector space, it has a linearly independent **extension**.

Theorem 1.1.2. *Let V be a vector space, let S be a nonempty subset of V such that $\operatorname{span}(S) \neq V$, and let v be a vector in $V \diagdown \operatorname{span}(S)$. Then S is linearly independent if and only if $S \cup \{v\}$ is linearly independent.*

Proof. ($\Rightarrow$): Suppose $av + b^1 s_1 + \cdots + b^k s_k = 0$ for distinct vectors $s_1, \dots, s_k$ in S and real numbers $a, b^1, \dots, b^k$. Then $a = 0$; for if not, then

$$v = -\left[\left(\frac{b^1}{a}\right) s_1 + \cdots + \left(\frac{b^k}{a}\right) s_k\right],$$

hence v is in $\operatorname{span}(V)$, which is a contradiction. Thus, $b^1 s_1 + \cdots + b^k s_k = 0$, and since S is linearly independent, we have $b^1 = \cdots = b^k = 0$.

($\Leftarrow$): As remarked above, any subset of a linearly independent set is linearly independent. $\square$

A (not necessarily finite) subset $\mathcal{H}$ of a vector space V is said to be an **unordered basis** for V if it spans V and is linearly independent.

Theorem 1.1.3. *If V is a vector space and $\mathcal{H}$ is an unordered basis for V, then each vector in V can be expressed uniquely (up to order of terms) as a linear combination of vectors in $\mathcal{H}$.*

Proof. Since $\mathcal{H}$ spans V, each vector in V can be expressed as a linear combination of vectors in $\mathcal{H}$. Suppose a vector v in V can be expressed as a linear combination in two ways. Let $h_1, \dots, h_k$ be the distinct vectors in the linear combinations. Then

$$v = a_1 h_1 + \cdots + a_k h_k \qquad \text{and} \qquad v = b_1 h_1 + \cdots + b_k h_k,$$

for some real numbers $a_1, \dots, a_k, b_1, \dots, b_k$, hence

$$(a_1 - b_1) h_1 + \cdots + (a_k - b_k) h_k = 0.$$

Since $\mathcal{H}$ is linearly independent, $a_i - b_i = 0$ for $i = 1, \dots, k$. $\square$

Theorem 1.1.4. *Let V be a vector space, and let S and T be nonempty subsets of V, where S is linearly independent, and T is finite and spans V. Then S is finite and $\operatorname{card}(S) \leq \operatorname{card}(T)$, where card denotes cardinality.*

Proof. Since S is linearly independent, it does not contain the zero vector. Let $\text{card}(T) = m$ and $T = \{t_1, \ldots, t_m\}$. We proceed in steps. For the first step, let s_1 be a vector in S. Since $V = \text{span}(T)$, s_1 is a linear combination of $t_1, \ldots, t_m$. Because s_1 is not the zero vector, at least one of the coefficients in the linear combination must be nonzero. Renumbering $t_1, \ldots, t_m$ if necessary, suppose it is the coefficient of t_1, and let $S_1 = \{s_1, t_2, \ldots, t_m\}$. Then t_1 can be expressed as a linear combination of the vectors in S_1, hence $V = \text{span}(S_1)$. For the second step, let s_2 be a vector in $S \setminus \{s_1\}$. Since $V = \text{span}(S_1)$, s_2 is a linear combination of $s_1, t_2, \ldots, t_m$. Because s_1, s_2 are linearly independent, at least one of the coefficients of $t_2, \ldots, t_m$ in the linear combination is nonzero. Renumbering $t_2, \ldots, t_m$ if necessary, suppose it is the coefficient of t_2, and let $S_2 = \{s_1, s_2, t_3, \ldots, t_m\}$. Then t_2 can be expressed as a linear combination of the vectors in S_2, hence $V = \text{span}(S_2)$. Proceeding in this way, after $k \leq m$ steps, we have a set $S_k = \{s_1, \ldots, s_k, t_{k+1}, \ldots, t_m\}$, with $V = \text{span}(S_k)$. Then $\text{card}(S) \leq \text{card}(T)$; for if not, at the mth step, we would have $S_m = \{s_1, \ldots, s_m\}$, with $V = \text{span}(S_m)$ and $S \setminus S_m$ nonempty. Then any vector in $S \setminus S_m$ could be expressed as a linear combination of vectors in S_m, which contradicts the assumption that S is linearly independent. □

We say that a vector space is **finite-dimensional** if it has a finite unordered basis. Finite-dimensional vector spaces have an associated invariant that, as we will see, largely characterizes them.

Theorem 1.1.5. *If V is a finite-dimensional vector space, then every unordered basis for V has the same (finite) number of vectors. This invariant, denoted by* $\dim(V)$*, is called the* ***dimension of V****.*

Proof. Let $\mathcal{H}$ and $\mathcal{F}$ be bases for V, with $\mathcal{F}$ finite. By Theorem 1.1.4, $\mathcal{H}$ is finite and $\text{card}(\mathcal{H}) \leq \text{card}(\mathcal{F})$. Then $\mathcal{H}$ is finite, so we use Theorem 1.1.4 again and obtain $\text{card}(\mathcal{F}) \leq \text{card}(\mathcal{H})$. Thus, $\text{card}(\mathcal{H}) = \text{card}(\mathcal{F})$. □

For completeness, we assign the zero vector space the dimension 0:

$$\dim(\{0\}) = 0.$$

Theorem 1.1.6. *If V is a vector space of dimension m, then:*
(a) *Every subset of V that spans V contains at least m vectors.*
(b) *Every linearly independent subset of V contains at most m vectors.*

Proof. (a): Let $\mathcal{H}$ be an unordered basis for V, and suppose T is a subset of V that spans V. The result is trivial if T is infinite, so assume otherwise. Then Theorem 1.1.4 and Theorem 1.1.5 give $m = \text{card}(\mathcal{H}) \leq \text{card}(T)$.

(b): Suppose S is a linearly independent subset of V. Then Theorem 1.1.4 and Theorem 1.1.5 yield $\text{card}(S) \leq \text{card}(\mathcal{H}) = m$. □

Theorem 1.1.7. *Let V be a vector space of dimension m, and let U be a subspace of V. Then:*
(a) *U is finite-dimensional and* $\dim(U) \leq \dim(V)$.

(b) *If* $\dim(U) = \dim(V)$, *then* $U = V$.
(c) *If* $\dim(U) < \dim(V)$, *then any unordered basis for* U *can be* ***extended*** *to an unordered basis for* V. *That is, given an unordered basis* $\{h_1, \ldots, h_k\}$ *for* U, *there are vectors* $h_{k+1}, \ldots, h_m$ *in* V *such that* $\{h_1, \ldots, h_k, h_{k+1}, \ldots, h_m\}$ *is an unordered basis for* V.

Proof. (a): We proceed in steps. For the first step, let u_1 be a vector in U. If $\mathrm{span}(\{u_1\}) = U$, we are done. If not, for the second step, let u_2 be a vector in $U \setminus \mathrm{span}(\{u_1\})$. It follows from Theorem 1.1.2 that u_1, u_2 are linearly independent. If $\mathrm{span}(\{u_1, u_2\}) = U$, we are done, and so on. By Theorem 1.1.6(b), this process ends after $k \leq m$ steps. Then $u_1, \ldots, u_k$ are linearly independent and span U, which is to say that $\{u_1, \ldots, u_k\}$ is an unordered basis for U.

(b): Let $\mathcal{H}$ and $\mathcal{F}$ be bases for U and V, respectively, and suppose $U \neq V$. Since $U = \mathrm{span}(\mathcal{H})$, there is a vector v in $V \setminus \mathrm{span}(\mathcal{H})$. By Theorem 1.1.2, $\mathcal{H} \cup \{v\}$ is linearly independent. We have from Theorem 1.1.5 that

$$\mathrm{card}(\mathcal{H} \cup \{v\}) > \mathrm{card}(\mathcal{H}) = \dim(U) = \dim(V) = \mathrm{card}(\mathcal{F}),$$

which contradicts Theorem 1.1.6(b).

(c): Given the unordered basis $\{h_1, \ldots, h_k\}$ for U, the algorithm described in part (a) can be used to find vectors $h_{k+1}, \ldots, h_m$ in V such that $\{h_1, \ldots, h_k, h_{k+1}, \ldots, h_m\}$ is an unordered basis for V. □

Throughout the remainder of Part I, unless stated otherwise, all vector spaces are finite-dimensional.

Let V be a vector space, and let $\{h_1, \ldots, h_m\}$ be an unordered basis for V. The m-tuple $(h_1, \ldots, h_m)$ is said to be an **ordered basis** for V, as is any m-tuple derived from $(h_1, \ldots, h_m)$ by permuting $h_1, \ldots, h_m$. For example, $(h_1, h_2, \ldots, h_m)$ and $(h_2, h_1, \ldots, h_m)$ are distinct ordered bases for V.

Example 1.1.8 ($\mathbb{R}^m$). Let e_i be the vector in $\mathbb{R}^m$ defined by

$$e_i = (0, \ldots, 0, 1, 0, \ldots, 0),$$

where 1 is in the ith position and 0s are elsewhere for $i = 1, \ldots, m$. For real numbers $a^1, \ldots, a^m$, we have

$$a^1 e_1 + \cdots + a^m e_m = (a^1, \ldots, a^m),$$

from which it follows that $e_1, \ldots, e_m$ span $\mathbb{R}^m$ and are linearly independent. We refer to $\{e_1, \ldots, e_m\}$ as the **standard unordered basis** for $\mathbb{R}^m$, and to $(e_1, \ldots, e_m)$ as the **standard ordered basis** for $\mathbb{R}^m$. Thus, not surprisingly, $\mathbb{R}^m$ has dimension m. ◊

Throughout the remainder of Part I, unless stated otherwise, all bases are ordered.

Accordingly, we now refer to $(e_1, \ldots, e_m)$ as the **standard basis** for $\mathbb{R}^m$.

Let V and W be vector spaces. A map $A : V \longrightarrow W$ is said to be **linear** if

$$A(cv + w) = cA(v) + A(w)$$

for all vectors v, w in V and all real numbers c. Thus, a linear map respects vector space structure. Suppose A is in fact a linear map. Given a basis $\mathcal{H} = (h_1, \ldots, h_m)$ for V, let us denote

$$\big(A(h_1), \ldots, A(h_m)\big) \qquad \text{by} \qquad A(\mathcal{H}).$$

We say that A is a **linear isomorphism**, and that V and W are **isomorphic**, if A is bijective. To illustrate, let x be an indeterminate, and let

$$\mathbb{P}_m = \{a_0 + a_1 x + \cdots + a_m x^m : a_0, \ldots, a_m \in \mathbb{R}\}$$

be the set of real polynomials of degree at most m. From the properties of polynomials, it is easily shown that $\mathbb{P}_m$ is a vector space of dimension $m+1$, and that the map $A : \mathbb{R}^{m+1} \longrightarrow \mathbb{P}_m$ given by $A(a_0, \ldots, a_m) = a_0 + a_1 x + \cdots + a_m x^m$ for all vectors $(a_0, \ldots, a_m)$ in $\mathbb{R}^{m+1}$ is a linear isomorphism. Following Section B.5, we denote the existence of an isomorphism by $\mathbb{R}^{m+1} \approx \mathbb{P}_m$.

Since a linear isomorphism is a bijective map, it has an inverse map. The next result shows that the inverse of a linear isomorphism is automatically a linear isomorphism.

Theorem 1.1.9. *If V and W are vector spaces and $A : V \longrightarrow W$ is a linear isomorphism, then $A^{-1} : W \longrightarrow V$ is a linear isomorphism.*

Proof. By assumption, A^{-1} is bijective. Let w_1, w_2 be vectors in W, and let c be a real number. Since A is bijective, there are unique vectors v_1, v_2 in V such that $A(v_1) = w_1$ and $A(v_2) = w_2$. Then

$$\begin{aligned} A^{-1}(cw_1 + w_2) &= A^{-1}\big(cA(v_1) + A(v_2)\big) = A^{-1}\big(A(cv_1 + v_2)\big) \\ &= cv_1 + v_2 = cA^{-1}(w_1) + A^{-1}(w_2). \end{aligned}$$

□

A linear map is completely determined by its values on a basis, as we now show.

Theorem 1.1.10. *Let V and W be vector spaces, let $\mathcal{H} = (h_1, \ldots, h_m)$ be a basis for V, and let $w_1, \ldots, w_m$ be vectors in W. Then there is a unique linear map $A : V \longrightarrow W$ such that $A(\mathcal{H}) = (w_1, \ldots, w_m)$.*

Proof. Uniqueness. Since $\mathcal{H}$ is a basis for V, for each vector v in V, there is a unique m-tuple $(a^1, \ldots, a^m)$ in $\mathbb{R}^m$ such that $v = a^1 h_1 + \cdots + a^m h_m$. Suppose $A : V \longrightarrow W$ is a linear map such that $A(\mathcal{H}) = (w_1, \ldots, w_m)$. Then

$$\begin{aligned} A(v) &= A(a^1 h_1 + \cdots + a^m h_m) \\ &= a^1 A(h_1) + \cdots + a^m A(h_m) \qquad (1.1.1) \\ &= a^1 w_1 + \cdots + a^m w_m, \end{aligned}$$

from which it follows that A is unique.

Existence. Let us define $A : V \longrightarrow W$ using (1.1.1) for all vectors v in V. The uniqueness of the m-tuple $(a^1, \ldots, a^m)$ ensures that A is well-defined. Clearly, $A(\mathcal{H}) = (w_1, \ldots, w_m)$. Let $u = b^1 h_1 + \cdots + b^m h_m$ be a vector in V, and let c be a real number. Then

$$cv + u = (ca^1 + b^1)h_1 + \cdots + (ca^m + b^m)h_m,$$

hence

$$\begin{aligned} A(cv + u) &= (ca^1 + b^1)A(h_1) + \cdots + (ca^m + b^m)A(h_m) \\ &= (ca^1 + b^1)w_1 + \cdots + (ca^m + b^m)w_m \\ &= c(a^1 w_1 + \cdots + a^m w_m) + (b^1 w_1 + \cdots + b^m w_m) \\ &= cA(v) + A(u). \end{aligned}$$

Thus, A is linear. □

From the point of view of linear structure, isomorphic vector spaces are indistinguishable. In fact, it is easily shown using Theorem 1.1.10 that all m-dimensional vector space are isomorphic. More than that, they are all isomorphic to $\mathbb{R}^m$. The isomorphism constructed with the help of Theorem 1.1.10 depends on the choice of bases for the vector spaces. However, we will see an instance in Section 1.2 where an isomorphism can be defined without having to resort to such an arbitrary choice.

Let V and W be vector spaces, and let $A : V \longrightarrow W$ be a linear map. The **kernel of *A*** is defined by

$$\ker(A) = \{v \in V : A(v) = 0\},$$

and the **image of *A*** by

$$\operatorname{im}(A) = \{A(v) \in W : v \in V\}.$$

It is easily shown that $\ker(A)$ is a subspace of V, and $\operatorname{im}(A)$ is a subspace of W. The **nullity of *A*** is defined by

$$\operatorname{null}(A) = \dim\bigl(\ker(A)\bigr),$$

and the **rank of *A*** by

$$\operatorname{rank}(A) = \dim\bigl(\operatorname{im}(A)\bigr).$$

The nullity and rank of a linear map satisfy an important identity.

Theorem 1.1.11 (Rank–Nullity Theorem). *If V and W are vector spaces and $A : V \longrightarrow W$ is a linear map, then*

$$\dim(V) = \operatorname{rank}(A) + \operatorname{null}(A). \tag{1.1.2}$$

Proof. By Theorem 1.1.7(c), any basis $(h_1,\ldots,h_k)$ for $\ker(A)$ can be extended to a basis $(h_1,\ldots,h_k,h_{k+1},\ldots,h_m)$ for V. We claim that $\big(A(h_{k+1}),\ldots,A(h_m)\big)$ is a basis for $\operatorname{im}(A)$. Let v be a vector in V. Since $\mathcal{H}$ spans V, we have $v = a^1h_1+\cdots+a^mh_m$ for some real numbers $a^1,\ldots,a^m$. Then

$$\begin{aligned} A(v) &= a^1A(h_1)+\cdots+a^kA(h_k)+a^{k+1}A(h_{k+1})+a^mA(h_m)\\ &= a^{k+1}A(h_{k+1})+a^mA(h_m), \end{aligned}$$

hence $A(h_{k+1}),\ldots,A(h_m)$ span $\operatorname{im}(A)$. Suppose

$$c^{k+1}A(h_{k+1})+\cdots+c^mA(h_m)=0$$

for some real numbers $c^{k+1},\ldots,c^m$. Then $A(c^{k+1}h_{k+1}+\cdots+c^mh_m)=0$, so $c^{k+1}h_{k+1}+\cdots+c^mh_m$ is in $\ker(V)$. Since $h_1,\ldots,h_k$ span $\ker(A)$, there are real numbers $b^1,\ldots,b^k$ such that

$$b^1h_1+\cdots+b^kh_k = c^{k+1}h_{k+1}+\cdots+c^mh_m,$$

hence

$$b^1h_1+\cdots+b^kh_k+(-c^{k+1})h_{k+1}+\cdots+(-c^m)h_m=0.$$

From the linear independence of $h_1,\ldots,h_k,h_{k+1},\ldots,h_m$, we have $c^{k+1}=\cdots=c^m=0$. Thus, $A(h_{k+1}),\ldots,A(h_m)$ are linearly independent. This proves the claim. It follows that

$$\begin{aligned} \operatorname{rank}(A) &= \dim\big(\operatorname{im}(A)\big) = m-k = \dim(V)-\dim\big(\ker(A)\big)\\ &= \dim(V)-\operatorname{null}(A). \end{aligned}$$

□

As an example of the rank–nullity identity, consider the linear map $A:\mathbb{R}^3\longrightarrow\mathbb{R}^2$ given by $A(x,y,z)=(x+y,0)$. Then

$$\ker(A)=\{(x,y,z)\in\mathbb{R}^3 : x+y=0\}$$

and

$$\operatorname{im}(A)=\{(x,y)\in\mathbb{R}^2 : y=0\}.$$

In geometric terms, $\ker(A)$ is a plane in $\mathbb{R}^3$ and $\operatorname{im}(A)$ is a line in $\mathbb{R}^2$. Thus, $\operatorname{null}(A)=2$ and $\operatorname{rank}(A)=1$, which agrees with Theorem 1.1.11.

In the notation of Theorem 1.1.11, we observe from (1.1.2) that $\operatorname{rank}(A)\le\dim(V)$. Thus, a linear map at best "preserves" dimension, but never increases it.

Theorem 1.1.12. *If V and W are vector spaces and $A:V\longrightarrow W$ is a linear map, then the following are equivalent:*
(a) $\operatorname{rank}(A)=\dim(V)$.
(b) $\operatorname{null}(A)=0$.
(c) $\ker(A)=\{0\}$.
(d) *A is injective.*

Proof. (a) ⇔ (b) ⇔ (c): By Theorem 1.1.11,

$$\begin{aligned} &\operatorname{rank}(A) = \dim(V) \\ &\quad \Leftrightarrow \quad \operatorname{null}(A) = 0 \\ &\quad \Leftrightarrow \quad \dim\big(\ker(A)\big) = 0 \\ &\quad \Leftrightarrow \quad \ker(A) = \{0\}. \end{aligned}$$

(c) ⇒ (d): For vectors v, w in V, we have

$$\begin{aligned} &A(v) = A(w) \\ &\quad \Leftrightarrow \quad A(v - w) = 0 \\ &\quad \Leftrightarrow \quad v - w \text{ is in } \ker(A) \\ &\quad \Rightarrow \quad v - w = 0. \end{aligned}$$

(d) ⇒ (c): Clearly, 0 is in $\ker(V)$. For a vector v in V, we have

$$\begin{aligned} &v \text{ is in } \ker(A) \\ &\quad \Leftrightarrow \quad A(v) = 0 \\ &\quad \Leftrightarrow \quad A(v) = A(0) \\ &\quad \Rightarrow \quad v = 0. \end{aligned}$$

□

Theorem 1.1.13. *Let V and W be vector spaces, let $\mathcal{H}$ be a basis for V, and let $A : V \longrightarrow W$ be a linear map. Then:*
(a) *A is a linear isomorphism if and only if $A(\mathcal{H})$ is a basis for W.*
(b) *If A is a linear isomorphism, then* $\dim(V) = \dim(W)$.

Proof. Let $\mathcal{H} = (h_1, \ldots, h_m)$.

(a)(⇒): Since A is surjective, for each vector w in W, there is a vector v in V such that $A(v) = w$. Let $v = a^1 h_1 + \cdots + a^m h_m$ for some real numbers $a^1, \ldots, a^m$. Then

$$w = A(v) = a^1 A(h_1) + \cdots + a^m A(h_m),$$

so $\mathcal{A}(\mathcal{H})$ spans W. Suppose $b^1 A(h_1) + \cdots + b^m A(h_m) = 0$ for some real numbers $b^1, \ldots, b^m$. Then $A(b^1 h_1 + \cdots + b^m h_m) = 0$, hence $b^1 h_1 + \cdots + b^m h_m$ is in $\ker(A)$. Since A is injective, it follows from Theorem 1.1.12 that $b^1 h_1 + \cdots + b^m h_m = 0$, hence $b^1 = \cdots = b^m = 0$. Thus, $\mathcal{A}(\mathcal{H})$ is linearly independent.

(a)(⇐): Let w be a vector in W. Since $A(\mathcal{H})$ spans W, we have $w = b^1 A(h_1) + \cdots + b^m A(h_m)$ for some real numbers $b^1, \ldots, b^m$. Then $w = A(b^1 h_1 + \cdots + b^m h_m)$, so A is surjective. Let $v = a^1 h_1 + \cdots + a^m h_m$ be a vector in $\ker(A)$. Then $0 = A(v) = a^1 A(h_1) + \cdots + a^m A(h_m)$. Since $A(\mathcal{H})$ is linearly independent, it follows that $a^1 = \cdots = a^m = 0$, so $v = 0$. Thus, $\ker(A) = \{0\}$. By Theorem 1.1.12, A is injective.

(b): This follows from part (a). □

We pause here to comment on the way proofs are presented when there is an equation or other type of display that stretches over several lines of text. The necessary justification for logical steps in such displays, whether it be equation numbers, theorem numbers, example numbers, and so on, are often provided in brackets at the end of corresponding lines. In order to economize on space, "[Theorem x.y.z]" and "[Example x.y.z]" are abbreviated to "[Th x.y.z]" and "[Ex x.y.z]". The proof of the next result illustrates these conventions.

Theorem 1.1.14. *If V and W are vector spaces of dimension m and $A : V \longrightarrow W$ is a linear map, then the following are equivalent:*
(a) *A is a linear isomorphism.*
(b) *A is injective.*
(c) *A is surjective.*
(d) $\mathrm{rank}(A) = m$.

Proof. (a) $\Rightarrow$ (b): This is true by definition.

(b) $\Leftrightarrow$ (c): By Theorem 1.1.11,

$$\dim(W) = \dim(V) = \mathrm{rank}(A) + \mathrm{null}(A) = \dim\big(\mathrm{im}(A)\big) + \mathrm{null}(A),$$

hence

$$\begin{aligned} & W = \mathrm{im}(A) \\ \Leftrightarrow \quad & \mathrm{null}(A) = 0 && \text{[Th 1.1.7(b)]} \\ \Leftrightarrow \quad & A \text{ is injective.} && \text{[Th 1.1.12]} \end{aligned}$$

(c) $\Rightarrow$ (a): Since A is surjective, we have from (b) $\Leftrightarrow$ (c) that A is also injective.

(d) $\Leftrightarrow$ (b): This follows from Theorem 1.1.12. □

Let V be a vector space, and let $U_1, \ldots, U_k$ be subspaces. The **sum of $\boldsymbol{U_1, \ldots, U_k}$** is denoted by $U_1 + \cdots + U_k$ and defined by

$$U_1 + \cdots + U_k = \{u_1 + \cdots + u_k : u_1 \in U_1, \ldots, u_k \in U_k\}.$$

For example, $\mathbb{R}(1,0) + \mathbb{R}(0,1) = \mathbb{R}^2$. It is easily shown that

$$U_1 + \cdots + U_k = \mathrm{span}(U_1 \cup \cdots \cup U_k),$$

from which it follows that $U_1 + \cdots + U_k$ is the smallest subspace of V containing each of $U_1, \ldots, U_k$, in the sense that any subspace containing each of $U_1, \ldots, U_k$ also contains $U_1 + \cdots + U_k$. We observe that

$$U_1 + \cdots + U_k + \{0\} = U_1 + \cdots + U_k,$$

which shows that adding the zero vector spaces does not change a sum. For vectors $v_1, \ldots, v_k$ in V, we have the following connection between spans and sums:

$$\mathrm{span}(\{v_1, \ldots, v_k\}) = \mathbb{R}v_1 + \cdots + \mathbb{R}v_k.$$