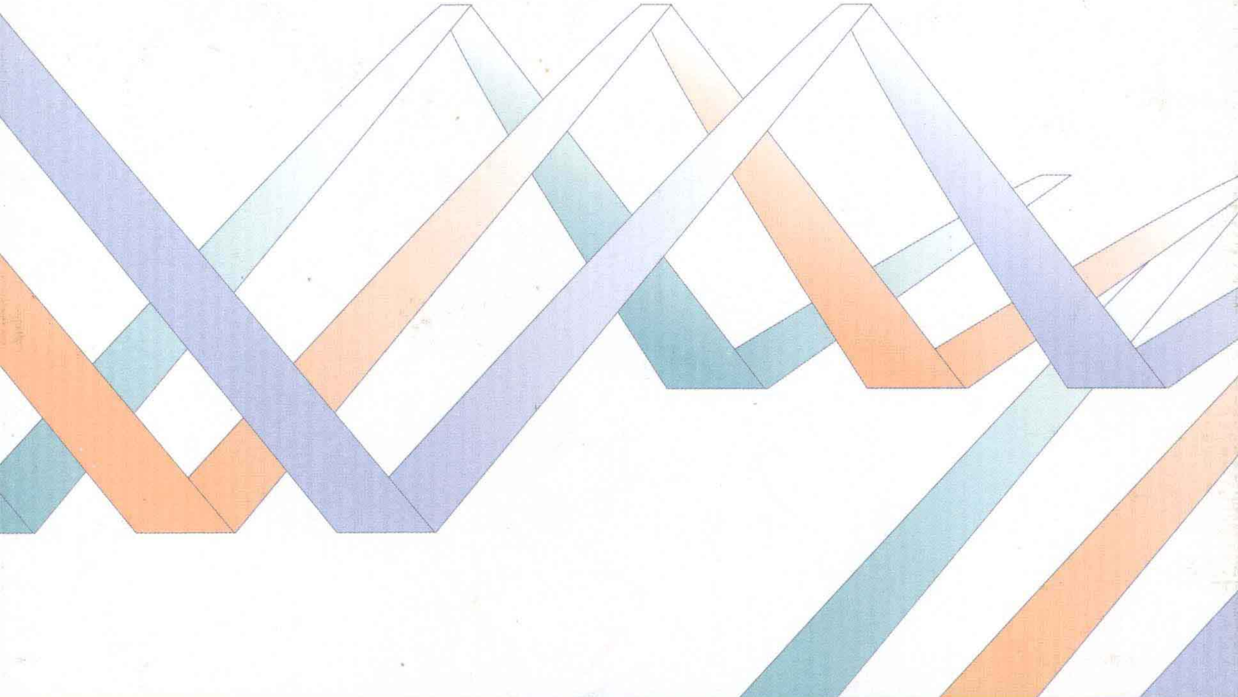
SMM 3
Surveys of Modern Mathematics



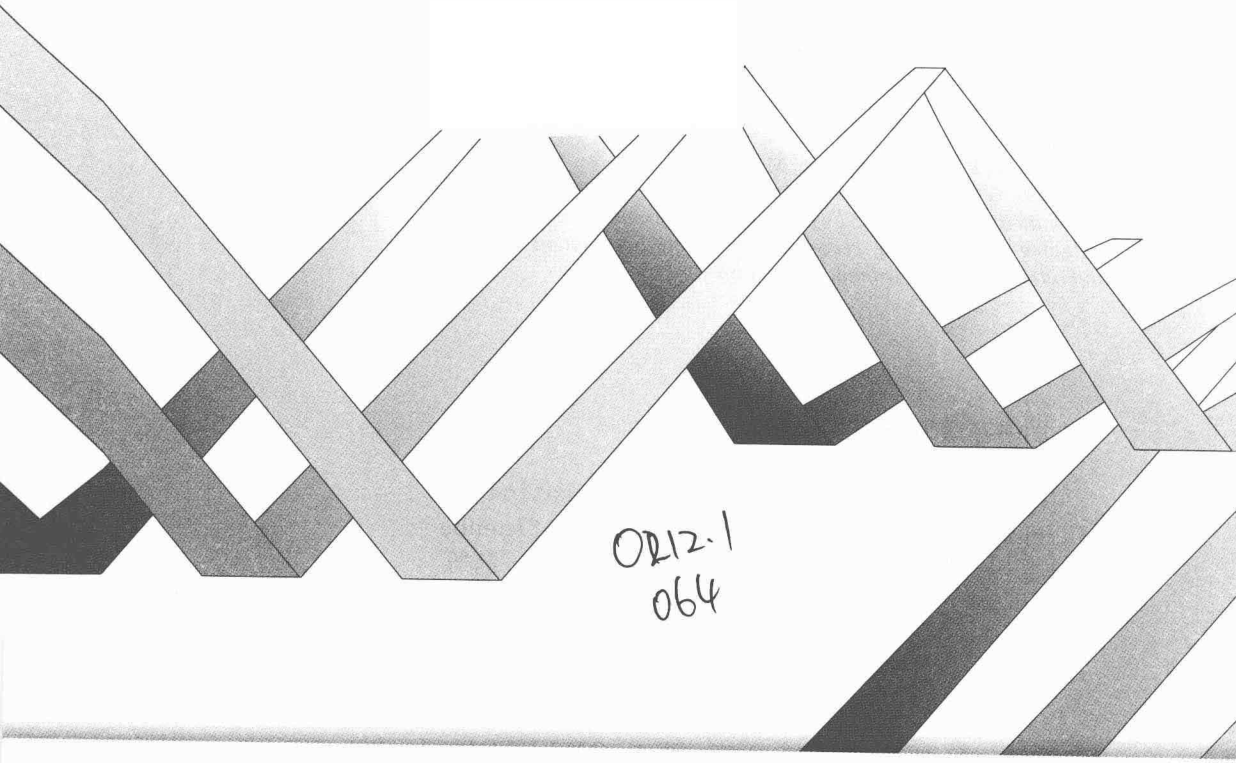
Application of Elementary Differential Geometry to Influence Analysis

微分几何在影响分析中的应用

Yat Sun Poon · Wai Yin Poon

高等教育出版社
HIGHER EDUCATION PRESS

SMM 3
Surveys of Modern Mathematics



Application of Elementary Differential Geometry to Influence Analysis

微分几何在影响分析中的应用

WEIFEN JIHE ZAI YINGXIANG FENXI ZHONG DE YINGYONG

Yat Sun Poon · Wai Yin Poon

高等教育出版社·北京
HIGHER EDUCATION PRESS BEIJING

International Press

图书在版编目 (CIP) 数据

微分几何在影响分析中的应用 = Application of Elementary Differential Geometry to Influence Analysis：英文 / 潘日新，潘伟贤编著. --北京：高等教育出版社, 2012.8
ISBN 978-7-04-035700-4

Ⅰ. ①微… Ⅱ. ①潘… ②潘… Ⅲ. ①微分几何-应用-统计分析-研究生-教材-英文 Ⅳ. ① O212.1②O186.1

中国版本图书馆 CIP 数据核字 (2012) 第 141353 号

策划编辑 王丽萍　　责任编辑 王丽萍　　封面设计 张申申　　版式设计 范晓红
责任校对 胡美萍　　责任印制 朱学忠

出版发行	高等教育出版社	咨询电话	400-810-0598
社　　址	北京市西城区德外大街4号	网　　址	http://www.hep.edu.cn
邮政编码	100120		http://www.hep.com.cn
印　　刷	涿州市星河印刷有限公司	网上订购	http://www.landraco.com
开　　本	787 mm×1092 mm　1/16		http://www.landraco.com.cn
印　　张	11.75	版　　次	2012年8月第1版
字　　数	250千字	印　　次	2012年8月第1次印刷
购书热线	010-58581118	定　　价	59.00元

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换

物 料 号　35700-00

Preface

The two authors of this book have different backgrounds within the mathematical sciences. After fifteen years of collaboration, they find it benefical to introduce their research areas jointly to students in the early stages of their research careers. Mathematics students will learn that mathematical concepts have an immediate impact on real life situations. Statistics students will appreciate that abstract mathematical concepts are accessible, relevant and valuable. In fact, an early draft of this book was used by one of the authors as lecture notes to run one year-long undergraduate research seminar for a mixture of mathematics and statistics undergraduate students in their senior years.

We hope that this book will encourage young scientists to develop an appreciation of inter-disciplinary research at a time when their research career is just beginning to emerge.

Differential geometry is a broad mathematical subject. Its global aspects are often presented only to mathematics majors or graduate students. However, its local aspects are the foundation of global issues and can be made accessible to all students with rigorous, multi-variable calculus training. To study local geometry, one studies graphs. It is a generalization of the freshman discussion on the concavity of a real-valued function of a single variable. By studying the geometry of graphs, one can proceed to learn global differential geometry as a mathematician.

One may also begin to apply the geometry of a graph to analyze functions arising from concrete problems, such as through statistics. Many statistical analyses involve a hypothesized model. Once a model is specified, the data collected are used to estimate the parameters that characterize the model. Further inference is affected by the hypothesized model and the data collected. Therefore, one must assess the influence based on the perturbation of various aspects of the model inputs. Perturbations can be represented by a set of perturbation parameters, and the function of these perturbation parameters becomes a mathematical object of interest. Therefore, one can apply geometric concepts to study this function and hence, deduce information on the perturbation.

Working through this process prompts theoretical, practical and technical issues. In particular, we must develop measures for the influence of individual perturbation parameters. We must also develop measures for the joint influence of any two, and then any groups of perturbation parameters. From single to multiple-parameter measures, the process is not a straightforward generalization of lower dimensional geometric problems to higher dimensional problems, because direct geometric generalization may not produce appropriate measures capable of serving the purposes of a statistical analysis. Instead, it is necessary to develop a set of meaningful measures that can constitute a well-structured system that enables further pursuit of relationships among the measures as well as the development of practical tools that facilitate interpretation and data analysis. Typical examples of such practical tools include the establishment of benchmarks for judging and interpreting measures and the search for alternative measures that reduce the computational burden. All such issues must be addressed using geometric techniques in the light of statistical considerations.

In Part I of this book, to fix conventions we recall basics of linear algebra, multi-variable calculus and Euclidean geometry in Chapters 1 and 2. In Chapter 3, we introduce the concept of normal sections, first fundamental forms and second fundamental forms. In Chapter 4, we introduce normal curvature and sectional curvatures. In Chapter 5, we study conformal transformations. This finishes our mathematical preparation.

In Part II, we first review some elementary statistics topics. In Chapters 6 and 7, we introduce basic concepts in relation to univariate distribution for discrete and continuous random variables, including the maximum likelihood estimation method. After generalizing these concepts to the bivariate and multivariate distributions in Chapter 8, we introduce simple linear regression in Chapter 9. Linear regression is the most popular statistical model, and is used as the key example in Part III of this book. To prepare for the illustration, some well-known topics in linear regression are discussed in Chapter 10.

In Part III, we apply the geometric concepts developed in Part I to the statistical issues and models articulated in Part II. The goal is to develop various measures generated by a local perturbation to assess the influence of the perturbation of model inputs. In Chapter 11, we develop the concept of likelihood displacement function. In Chapter 12, we apply the tools developed in Chapters 3 and 4 to analyze the likelihood displacement function. In the process of this development, different or mingled perspectives must be clarified. We use the regression model with some of the most well-known perspectives described in Chapter 10 for illustration. The relations among various measures that are generated in the process must also be analyzed. This is done in Chapter 13. Finally, in Chapter 14, we analyze the modification of perturbations using Chapter 5 as the theoretical foundation.

As we develop our presentation, we often encounter tedious but necessary computations. For the sake of completeness, we have placed computations in the Appendices.

Graduate students in mathematics may choose to begin reading this book starting with Part II. Likewise, graduate students in statistics may skip Part II. However, undergraduate students from both disciplines will benefit from reading this book in its entirety.

Yat-Sun Poon
Wai-Yin Poon
May, 2012

Contents

Part III Local Influence Analysis

Part I

Geometry

Chapter 1

Preliminaries

1.1 Linear algebra

In this section, we recall several pre-requisite facts of linear algebra on $\mathbb{R}^n$ and the maps between them.

1.1.1 Vectors and matrices

Here we review several linear algebra facts, to set up our conventions and provide a quick reminder of elementary materials.

We denote the n-dimensional real vector space by $\mathbb{R}^n$, and treat a point in $\mathbb{R}^n$ as a column vector. A vector $\mathbf{x}$ in $\mathbb{R}^n$ is given in terms of its coordinates:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \tag{1.1}$$

where $x_i \in \mathbb{R}$ for $1 \leqslant i \leqslant n$. Given $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$, the *vector addition* $\mathbf{x} + \mathbf{y}$ is the vector whose j-th coordinate is $x_j + y_j$ for all j from 1 to n. If λ is a real number, the *scalar multiplication* of λ on $\mathbf{x}$, denoted by $\lambda\mathbf{x}$ and $\mathbf{x}\lambda$, represents the vector whose j-th coordinate is λx_j. The vector whose coordinates are all equal to zero is called the zero-vector and is denoted by $\mathbf{0}$.

A *linear combination* of a collection of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ in $\mathbb{R}^n$ is a vector of the form $\lambda_1\mathbf{v}_1 + \cdots + \lambda_k\mathbf{v}_k$ for some real numbers $\lambda_1, \ldots, \lambda_k$. Note that there is not any restriction on the number k relative to the dimension n.

The *linear span* of a collection of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ in $\mathbb{R}^n$ is the set of all possible linear combinations for those vectors. If the linear span is equal to $\mathbb{R}^n$, we say that this collection spans $\mathbb{R}^n$.

A collection of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ in $\mathbb{R}^n$ is *linearly independent* if $\lambda_1 \mathbf{v}_1 + \cdots + \lambda_k \mathbf{v}_k = \mathbf{0}$ implies that each λ_j is equal to zero. If this collection of vectors spans $\mathbb{R}^n$, then it is called a *basis* for $\mathbb{R}^n$. In such a case, it is necessary that $k = n$.

We use $\mathbf{e}_j$ to denote the vector whose coordinates are all equal to zero except that its j-th coordinate is equal to 1. This collection of vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ spans $\mathbb{R}^n$ because if $\mathbf{x}$ is given as (1.1), then

$$\begin{aligned}\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} &= x_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \cdots + x_n \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \\ &= x_1 \mathbf{e}_1 + \cdots + x_n \mathbf{e}_n .\end{aligned}$$

It is also apparent that if $\mathbf{x} = \mathbf{0}$, then each x_j is equal to zero. Therefore, $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a basis for $\mathbb{R}^n$. We consider this particular basis as the *standard basis* for the vector space $\mathbb{R}^n$. The individual vectors are identified as *basic vectors.*

A map $\mathbf{A} : \mathbb{R}^n \to \mathbb{R}^m$ is *linear* if

$$\mathbf{A}(\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2) = \lambda_1 \mathbf{A}(\mathbf{v}_1) + \lambda_2 \mathbf{A}(\mathbf{v}_2)$$

for all real numbers λ_1, λ_2, and vectors $\mathbf{v}_1, \mathbf{v}_2$ in $\mathbb{R}^n$. Given an $m \times n$-matrix $\mathbf{A}$, a matrix multiplication from the left on a column vector defines a linear map. To be precise,

$$\mathbf{A}(\mathbf{x}) = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \tag{1.2}$$

where each entry a_{ij} is a real number. In this case, we also use matrix multiplication notation $\mathbf{Ax}$ to represent the map $\mathbf{A}(x)$. This means that if

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$$

are the coordinates of a point $\mathbf{y}$ in $\mathbb{R}^m$, then $\mathbf{Ax} = \mathbf{y}$ if and only if for all k in $\{1, \ldots, m\}$,

$$y_k = \sum_{j=1}^{n} a_{kj} x_j . \tag{1.3}$$

Because $\mathbf{Ae}_j$ is the j-th column of the matrix $\mathbf{A}$, one may re-write the above expression for the matrix $\mathbf{A}$ in terms of the result of its linear action on $\mathbf{e}_j$, i.e.,

$$\mathbf{A} = (\mathbf{Ae}_1, \ldots, \mathbf{Ae}_n) . \tag{1.4}$$

On the other hand, given the standard basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ and an arbitrary linear map L, by taking the coordinates of $L(\mathbf{e}_j)$ as the j-th column of a matrix the linear map L is uniquely represented by a matrix.

The *transpose* of an $m \times n$-matrix $\mathbf{A}^T$ is an $n \times m$-matrix obtained by interchanging the rows with columns in $\mathbf{A}$. To be precise, if $\mathbf{A}$ is given as in (1.2), then

$$\mathbf{A}^T = \begin{pmatrix} a_{11} & \ldots & a_{m1} \\ \vdots & & \vdots \\ a_{1n} & \ldots & a_{nm} \end{pmatrix} \tag{1.5}$$

For example, $\mathbf{x} = (x_1, \ldots, x_n)^T$ and $\mathbf{0} = (0, \ldots, 0)^T$.

An $n \times n$-matrix is also called a *square matrix.* A square matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^T$. In terms of its entries, $a_{ij} = a_{ji}$ for all $1 \leqslant i, j \leqslant n$. A square matrix $\mathbf{A}$ is *diagonal* if $a_{ij} = 0$ for all $i \neq j$. In such case, we will also denote $\mathbf{A}$ by its diagonal terms as $\mathrm{Diag}(a_{11}, \ldots, a_{nn})$. The square matrix $\mathrm{Diag}(1, \ldots, 1)$ is also denoted by $\mathbf{I}$, which is also known as the *identity matrix.*

Exercise 1.1.1 Suppose $\mathbf{A}$ is an $m \times n$-matrix and $\mathbf{B}$ is an $n \times k$-matrix, then $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$.

1.1.2 Symmetric bilinear forms

A *bilinear form* of the vector space $\mathbb{R}^n$ is a function g that associated with each pair of vectors $\mathbf{x}$ and $\mathbf{y}$, is real number $g(\mathbf{x}, \mathbf{y})$, which satisfies the following properties for all real numbers λ_1 and λ_2 and all vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_1$ and $\mathbf{y}_2$ in $\mathbb{R}^n$,

$$\begin{aligned} g(\lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2, \mathbf{y}) &= \lambda_1 g(\mathbf{x}_1, \mathbf{y}) + \lambda_2 g(\mathbf{x}_2, \mathbf{y}); \\ g(\mathbf{x}, \lambda_1\mathbf{y}_1 + \lambda_2\mathbf{y}_2) &= \lambda_1 g(\mathbf{x}, \mathbf{y}_1) + \lambda_2 g(\mathbf{x}, \mathbf{y}_2). \end{aligned}$$

A bilinear form is symmetric if $g(\mathbf{x}, \mathbf{y}) = g(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$.

Example 1.1.2 (Dot product) For all $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$, define

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T\mathbf{y} = \sum_{j=1}^{n} x_j y_j. \tag{1.6}$$

Then $\langle \mathbf{x}, \mathbf{y} \rangle$ is the usual dot product between the vectors $\mathbf{x}$ and $\mathbf{y}$.

Example 1.1.3 (Symmetric matrices) Given an $n \times n$-matrix $\mathbf{A}$, define a map from the product $g_{\mathbf{A}} : \mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}$ by

$$g_{\mathbf{A}}(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T\mathbf{A}\mathbf{y}. \tag{1.7}$$

The map $g_{\mathbf{A}}$ is a bilinear map associated with the matrix $\mathbf{A}$. If the matrix $\mathbf{A}$ is symmetric,

$$g_{\mathbf{A}}(\mathbf{x},\mathbf{y}) = \mathbf{x}^T\mathbf{A}\mathbf{y} = (\mathbf{x}^T\mathbf{A}\mathbf{y})^T = \mathbf{y}^T\mathbf{A}^T(\mathbf{x}^T)^T = \mathbf{y}^T\mathbf{A}\mathbf{x} = g_{\mathbf{A}}(\mathbf{y},\mathbf{x}). \quad (1.8)$$

This means that the bilinear map $g_{\mathbf{A}}$ associated with the matrix $\mathbf{A}$ is symmetric.

In particular, the dot product is the bilinear map associated with the identity matrix.

1.1.3 Vector subspaces

A subset V of $\mathbb{R}^m$ is a *vector subspace.* If for any two elements $\mathbf{v}_1$ and $\mathbf{v}_2$ in V and any two real numbers λ_1 and λ_2, $\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2$ is again an element in V. Because one may choose $\lambda_1 = 0$ and $\lambda_2 = 0$, a vector subspace necessarily contains the zero vector $\mathbf{0}$ of $\mathbb{R}^m$. The *dimension* of a vector subspace V is the number of independent vectors needed to span V.

For example, the set

$$V = \{\mathbf{x} \in \mathbb{R}^{n+1} : x_{n+1} = 0\} \quad (1.9)$$

is a vector subspace. One may consider this vector subspace as a copy of $\mathbb{R}^n$. It is an n-dimensional vector subspace in $\mathbb{R}^{n+1}$.

To be precise, consider a linear map $\mathbf{A} : \mathbb{R}^n \to \mathbb{R}^{n+1}$ defined by

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}. \quad (1.10)$$

As the image of $\mathbb{R}^n$ is equal to V, as defined above, and $\mathbf{A}\mathbf{x}_1 = \mathbf{A}\mathbf{x}_2$ if and only if $\mathbf{x}_1 = \mathbf{x}_2$, the map $\mathbf{A}$ identifies $\mathbb{R}^n$ to the vector subspace V in both one-to-one and onto fashions. We address the map $\mathbf{A}$ in (1.10) along with the subspace V the standard embedding of $\mathbb{R}^n$ in $\mathbb{R}^{n+1}$.

More generally, suppose that $\mathbf{A} : \mathbb{R}^m \to \mathbb{R}^n$ is a linear map. Consider the image and kernel of $\mathbf{A}$.

$$\text{Image}\mathbf{A} = \{\mathbf{A}\mathbf{v} \in \mathbb{R}^n : \mathbf{v} \in \mathbb{R}^m\}, \quad \ker\mathbf{A} = \{\mathbf{v} \in \mathbb{R}^m : \mathbf{A}\mathbf{v} = 0\}.$$

This is an elementary exercise to show that Image$\mathbf{A}$ is a vector subspace of $\mathbb{R}^n$ and ker$\mathbf{A}$ is a vector subspace of $\mathbb{R}^m$.

The *rank* of an $m \times n$-matrix $\mathbf{A}$ is the dimension of vector subspace Image$\mathbf{A}$ in $\mathbb{R}^m$.

Another kind of vector spaces often encountered can be constructed as follows. Let $\mathbf{n}$ be a non-zero vector in $\mathbb{R}^{n+1}$. Define

$$V^{\perp}(\mathbf{n}) = \{\mathbf{v} \in \mathbb{R}^n : \langle \mathbf{v}, \mathbf{n} \rangle = 0\}. \tag{1.11}$$

Due to the linearity of dot product, for any real numbers λ_1 and λ_2 and vectors $\mathbf{v}_1$ and $\mathbf{v}_2$,

$$\langle \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2, \mathbf{n} \rangle = \lambda_1 \langle \mathbf{v}_1, \mathbf{n} \rangle + \lambda_2 \langle \mathbf{v}_2, \mathbf{n} \rangle.$$

Therefore, if $\mathbf{v}_1$ and $\mathbf{v}_2$ are in the set $V^{\perp}(\mathbf{n})$, then $\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2$ is also contained in the set $V^{\perp}(\mathbf{n})$. This shows that $V^{\perp}(\mathbf{n})$ is a vector subspace of $\mathbb{R}^n$. Because it is the set of vectors orthogonal to the vector $\mathbf{n}$, it is called the *orthogonal complement* of $\mathbf{n}$. The vector $\mathbf{n}$ is called a *normal vector* of the subspace $V^{\perp}(\mathbf{n})$.

1.1.4 Linear maps from $\mathbb{R}^n$ to $\mathbb{R}^n$

A non-zero vector $\mathbf{x}$ is an *eigenvector* of the square matrix $\mathbf{A}$ if and only if there exists a real number λ, such that $\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$. The number λ is called an *eigenvalue* of the matrix $\mathbf{A}$. The set

$$V_\lambda = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \lambda \mathbf{x}\}$$

is called the eigenspace of the matrix $\mathbf{A}$. All eigenspaces are *invariant subspaces* of $\mathbf{A}$, i.e., whenever $\mathbf{v}$ is in V, $\mathbf{A}\mathbf{v}$ is in V.

Not every square matrix has eigenvalus. For example, the matrix

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

is a rotation of 90° on the two-dimensional plane. It does not leave any one-dimensional subspace invariant. However, we have the following.

Theorem 1.1.4 *Suppose that* $\mathbf{A} : \mathbb{R}^n \to \mathbb{R}^n$ *is a symmetric matrix, then there is a basis* $\{v_1, \ldots, v_n\}$ *for* $\mathbb{R}^n$*, such that each vector* v_j *in the basis is an eigenvector.*

A linear transformation $\mathbf{A}$ on $\mathbb{R}^n$ is *invertible* if there exists another linear transformation $\mathbf{B}$ on $\mathbb{R}^n$ such that $\mathbf{A}\mathbf{B}\mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^n$. In terms of matrices, it is equivalent to $\mathbf{A}\mathbf{B} = \mathbf{I}$. The matrix $\mathbf{B}$ is called the *inverse* of $\mathbf{A}$. It is uniquely determined by $\mathbf{A}$ and is denoted by $\mathbf{A}^{-1}$. It can also be proved that the identity $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ is equivalent to $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$.