

Article

Not peer-reviewed version

Differential Topology and Matrix Analysis: An Overview

[Petko H. Petkov](#)*

Posted Date: 23 January 2026

doi: 10.20944/preprints202601.1792.v1

Keywords: matrix computations; numerical analysis; smooth manifolds; smooth maps; singularities; conditioning; probabilistic bounds; regularization



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Differential Topology and Matrix Analysis: An Overview

Petko H. Petkov 

Department of Engineering Sciences, Bulgarian Academy of Sciences, Sofia 1040, Bulgaria; php@tu-sofia.bg

Abstract

The present overview aims to illustrate the application of differential topology methods to some important problems in matrix analysis. In particular, it focuses on the use of smooth manifolds and smooth mappings to study fundamental issues such as the determination of matrix rank and the computation of the Jordan form in presence of uncertainties. Various aspects of numerical matrix analysis are discussed, including the genericity of matrix problems, characterization of singular sets in the parameter space, the distance to ill-posed problems and its relation to problem conditioning. The paper also addresses the conditioning of matrix problems in both deterministic and probabilistic settings and the regularization of ill-posed matrix problems. Several examples are provided to illustrate these concepts and their practical relevance.

Keywords: matrix computations; numerical analysis; smooth manifolds; smooth maps; singularities; conditioning; probabilistic bounds; regularization

MSC: 53Z50; 53-01; 57R45; 58K05; 65F15; 65F22; 22E70

1. Introduction

Contemporary mathematical models used across the natural sciences are typically not single models corresponding to fixed parameter values; rather, they represent families of models that vary as parameters change within certain bounds. Such models arise in physics, chemistry, biology, control theory, statistics, aerodynamics, hydrodynamics, the social sciences, and several other disciplines. Studying these models using traditional methods is challenging, because the properties of a family of models cannot be described as continuous functions of the parameters if one considers only individual models corresponding to fixed parameter values. This difficulty necessitates the use of families of models whose descriptions and properties depend smoothly on the parameters. Consequently, in recent years, specialists from various scientific fields have shown growing interest in methods of differential topology, whose objects of study are smooth manifolds and smooth mappings. Representing mathematical models in the natural sciences as smooth manifolds, with dimension determined by the number of independent parameters, allows the essential properties of families of models to be expressed as smooth functions of these parameters, thereby greatly facilitating the solution of the corresponding mathematical problems. In this connection, it is appropriate to cite the German mathematician and philosopher Hermann Weyl, who wrote [1], p. 90: *Topology has the peculiarity that questions belonging to its domain may under certain circumstances be decidable, even though the continua to which they are addressed may not be given exactly but only vaguely, as is always the case in reality.*

The questions addressed by differential topology are global in nature, as they concern the manifold as a whole. Differential topology combines the study of qualitative properties of sets in spaces of arbitrary dimension, which is the domain of topology, with the methods of classical analysis, which enable quantitative analysis under small parametric variations. In this regard, it is appropriate to quote the words of the American mathematician Marston Morse [2], Foreword:

Any problem which is nonlinear in character, which involves more than one coordinate system or variable, or

whose structure is initially defined in the large, is likely to require considerations of topology and group theory in order to arrive at its meaning and its solution. In the solution of such problems classical analysis will frequently appear as an instrument in the small, integrated over the whole problem with the aid of group theory or topology.

Differential topology is a broad mathematical discipline whose primary goal is the study and characterization of the global properties of manifolds. A central theme in this field is the transition from local to global properties: many concepts in differential topology can be understood by examining how local behavior extends to the global structure. Another fundamental notion is manifold transversality, which describes the manner in which two manifolds intersect and provides a framework for understanding generic intersections and their stability.

The present overview aims to illustrate the application of methods from differential topology to the solution of several important mathematical problems arising across the natural and applied sciences. In particular, it focuses on the use of smooth manifolds and smooth mappings in matrix analysis, a field with wide-ranging applications in physics, engineering, biology, and beyond. The motivation for this overview arises from the absence of a single comprehensive reference on this subject, as the relevant material is currently scattered across numerous books and research articles.

It should be noted that some mathematical rigor has been deliberately relaxed to make the exposition accessible to specialists from different disciplines.

The overview is organized into six sections.

In Section 2, we present the basic concepts from differential topology necessary for the subsequent discussion. We briefly consider smooth manifolds and smooth maps between them, including the differential of a map and singular points of varieties. Some fundamental facts about Lie groups and matrix groups are also included.

Section 3 is devoted to the geometry of matrix spaces. We examine important characteristics of problems in this space, such as genericity and well-posedness. Condition numbers of matrix problems are discussed in detail to demonstrate the connection between the distance to ill-posed problems and problem conditioning. We also present results on the probabilistic distribution of matrix condition numbers obtained using methods from differential topology.

The important problem of matrix rank is considered in Section 4. We study the orbits of matrices with different ranks and show how small matrix perturbations can move a matrix to an orbit with lower codimension. The problem of determining the numerical rank of a matrix in the presence of uncertainties is also discussed.

Another fundamental problem in matrix analysis—the determination of the Jordan form of a matrix—is addressed in Section 5. We consider orbits and bundles of matrices with fixed Jordan form and investigate their generic properties. The reduction to the “true” Jordan form is described as an ill-posed problem, whose solution can be obtained via regularization methods. This leads to the concept of the numerical Jordan form, which is defined using the tools of differential topology.

In Section 6, we study matrices depending on parameters. It is shown that smooth properties of such matrices can be determined using versal deformations. Several examples of bifurcation diagrams are provided to illustrate how the Jordan form of a matrix depends on the varying parameters.

All computations in this paper were performed using MATLAB[®] Version 9.9 (R2020b) [3], employing IEEE double precision arithmetic with a unit roundoff $u \approx 1.11 \times 10^{-16}$.

2. A Glimpse into Differential Topology

The presentation in this section follows the classic textbooks by Guillemin and Pollack [4], Lee [5,6], and Arnold [7,8], as well as the books by Tu [9] and by Burns and Gidea [10], which are written in a language accessible to non-mathematicians. Excellent introductions to manifold theory for non-specialists include the books by Milnor [11] and Wallace [12]. One of the most authoritative sources in this field is Hirsch’s book [13]; however, reading it requires a very strong mathematical background. Applications of manifolds in mechanics are discussed in depth in [14].

2.1. Smooth Manifolds

A manifold is a multidimensional generalization of the concepts of a line and a surface, without singular points. When studying manifolds, the notion of dimension plays a central role. Generally speaking, the dimension is the number of independent quantities (or parameters) required to specify a point on the manifold. Manifolds of dimension one are lines and curves, while manifolds of dimension two are surfaces. Typical examples of two-dimensional manifolds include planes and spheres, as well as other familiar surfaces such as cylinders, ellipsoids, paraboloids, and tori. A key feature of these examples is that an n -dimensional manifold “looks” locally like \mathbb{R}^n : every point of a manifold \mathcal{M} has a neighborhood that is topologically equivalent to an open subset of \mathbb{R}^n . Thus, in one-dimensional manifolds each point has a neighborhood resembling a line segment; in two-dimensional manifolds, each point has a neighborhood resembling an open disk; and in three-dimensional manifolds, a neighborhood resembling an open ball.

In this sense, manifolds are sets in which the neighborhood of every point has the same local topological structure as the n -dimensional Euclidean space.

We note that the concept of dimension, as used in the characterization of manifolds, belongs to the most fundamental ideas in mathematics. An excellent overview of the significance of this concept in geometry and algebra is given by Manin in [15].

In Figure 1 we illustrate the decomposition of three-dimensional Euclidean space into layers (or strata) of manifolds defined by the equation $x^2 + y^2 - z^2 = C$ for different values of C . An essential feature of this decomposition is that the individual layers do not intersect. Note that the innermost layer (the two opposite cones with a common vertex at the origin) is not smooth; rather, it is an algebraic manifold, since the vertex is a singular point (see Section 2.7).

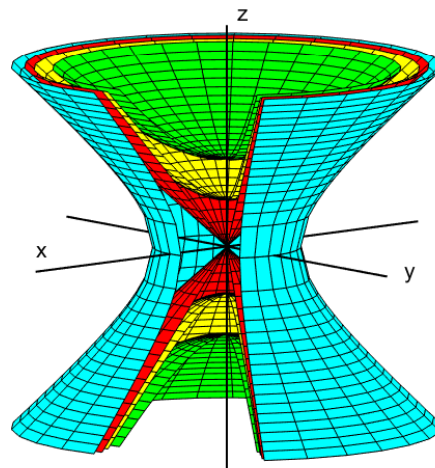


Figure 1. Manifolds in the 3-dimensional space.

Definition 1. Two subsets of Euclidean spaces $U \subseteq \mathbb{R}^k$, $V \subseteq \mathbb{R}^n$ are **topologically equivalent**, or **homeomorphic** (from the Greek word meaning “similar form”), if there exists a one-to-one correspondence $\varphi: U \rightarrow V$, such that both φ and its inverse are continuous. Such a correspondence is called a **homeomorphism**.

Based on these considerations, a provisional definition of a topological manifold can be given. We can consider **n -dimensional manifold** as a subset of some Euclidean space \mathbb{R}^k that is locally Euclidean of dimension n , i.e., every point of \mathcal{M} has a neighborhood in \mathcal{M} that is homeomorphic to a ball in \mathbb{R}^k .

For example, every one-dimensional manifold is homeomorphic to either a line or a circle.

Definition 2. A topological space \mathcal{M} is called an **n -dimensional topological manifold**, if it is locally homeomorphic to \mathbb{R}^n .

Every topological manifold \mathcal{M} is a Hausdorff space: for every pair of distinct points $p, q \in \mathcal{M}$, there exist disjoint open subsets $U, V \subseteq \mathcal{M}$ such that $p \in U$ and $q \in V$.

In most cases, the analysis of manifolds cannot be performed directly on the manifold itself. Instead, it is necessary to describe the manifold unambiguously in an appropriate coordinate space and to apply analytical methods to this representation. For this purpose, coordinate charts and a manifold atlas are employed.

An *open chart* of \mathcal{M} is defined as a pair (U, φ) , where U is an open subset of the space \mathcal{M} , and φ a homeomorphism from U onto an open subset of the coordinate space \mathbb{R}^n . To each point $p \in U$ there corresponds, in a one-to-one manner, an n -tuple of numbers

$$\varphi: p \in U \mapsto (x^1(p), \dots, x^n(p)) \in \varphi(U),$$

which are called its *local coordinates* (Figure 2).

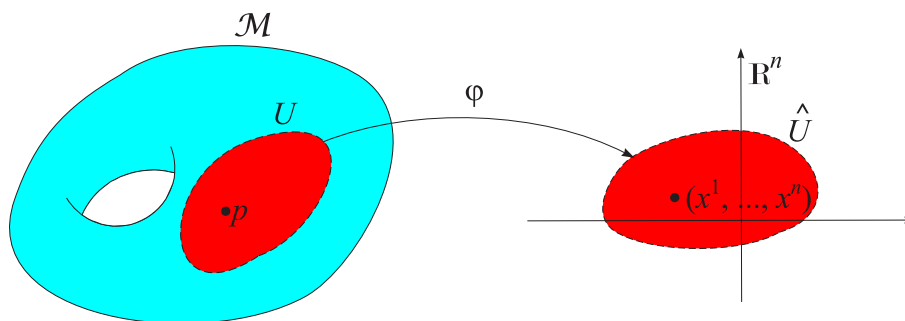


Figure 2. Open chart.

On the basis of the concept of an open chart, a rigorous definition of a topological manifold can be introduced.

Let A be a finite or countable set. A topological space \mathcal{M} is called an *n -dimensional topological manifold* if there exists a collection of open charts $(U_\alpha, \varphi_\alpha)$, $\alpha \in A$, such that:

1. For each α of the set A , $\varphi_\alpha(U_\alpha)$ is an open subset of \mathbb{R}^n .
2. $\mathcal{M} = \bigcup_{\alpha \in A} U_\alpha$.

Such a collection of charts is called an *atlas* of the topological manifold (Figure 3). The atlas of \mathcal{M} is denoted by $\mathcal{A}(\mathcal{M})$.

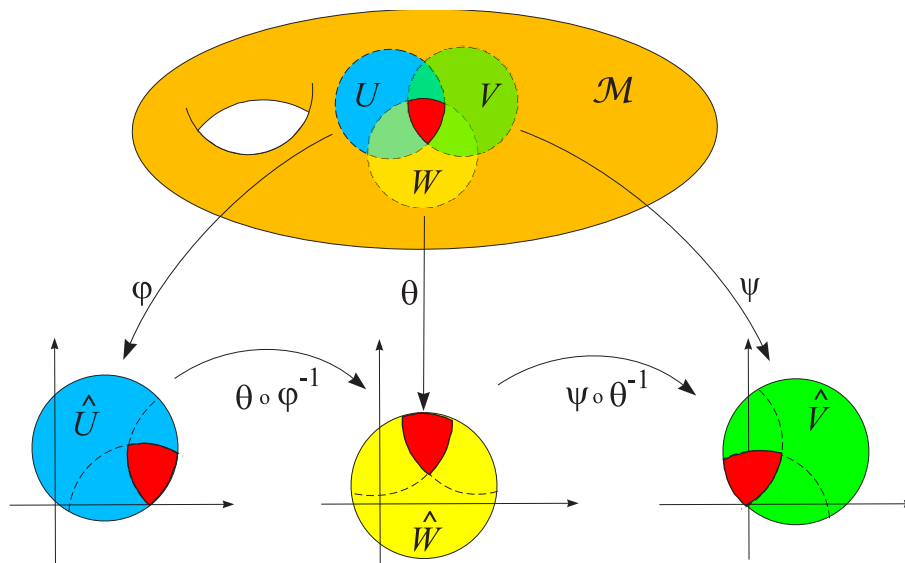


Figure 3. An atlas of a manifold.

In the general case, each chart (U_i, φ_i) is obtained using a different mapping φ_i associated with the subset U_i . In this way, one can study complex manifolds composed of several subsets with different properties. This constitutes an important advantage of manifolds over simpler topological objects which consist of a single set with fixed properties and are described by only one chart.

Example 1. (a) The coordinate space \mathbb{R}^n is an n -dimensional topological manifold: its atlas consists of a single open chart $(\mathbb{R}^n, i_{\mathbb{R}^n})$, where $i_{\mathbb{R}^n}: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the identity map.

(b) **(Atlas of the two-dimensional sphere).**

Let $N = (0, \dots, 0, 1) \in \mathbb{S}^k$ denote the **north pole** of the k -dimensional sphere, and let $S = (0, \dots, 0, -1)$ denote its **south pole**.

The **stereographic projection** φ_N from $U_1 = \mathbb{S}^k \setminus \{N\}$ onto \mathbb{R}^k is the mapping that sends a point p to the point where the line through N and p intersects the subspace of \mathbb{R}^{k+1} defined by $x^{k+1} = 0$ (the projection plane). See Figure 4 for the case $k = 2$. It is a smooth, bijective map from the entire sphere, except for the projection point, onto the whole plane. Stereographic projection provides a way to represent the sphere by a plane, but it can also be used for other curved surfaces, such as deformed spheres and hyperboloids.

The map $\varphi_N: \mathbb{S}^k \setminus \{N\} \rightarrow \mathbb{R}^k$ is given by the formula

$$(x^1, \dots, x^{k+1}) \mapsto \frac{1}{1 - x^{k+1}} (x^1, \dots, x^k).$$

Analogously, the projection φ_S from $U_2 = \mathbb{S}^k \setminus \{S\}$ onto \mathbb{R}^k is defined by

$$(x^1, \dots, x^{k+1}) \mapsto \frac{1}{1 + x^{k+1}} (x^1, \dots, x^k).$$

These projections are homeomorphisms from U_1 and U_2 onto \mathbb{R}^2 .

Let us define on the sphere $\mathbb{S}^2 \subset \mathbb{R}^3$ an atlas consisting of two charts, using the stereographic projections

$$\varphi_N(x) = \left(\frac{x^1}{1 - x_3}, \frac{x^2}{1 - x_3} \right), \quad \varphi_S(x) = \left(\frac{x^1}{1 + x_3}, \frac{x^2}{1 + x_3} \right).$$

In this case, the open sets $\varphi_N(U_1)$ and $\varphi_S(U_2)$, where $U_1 = \mathbb{S}^2 \setminus \{N\}$ and $U_2 = \mathbb{S}^2 \setminus \{S\}$, are planes in \mathbb{R}^3 that are tangent to the sphere at the points S and N , respectively.

The atlas of the sphere is shown in Figure 5. The family of circles lying on the sphere and tangent at the point N is mapped in the lower chart to a family of parallel straight lines, while in the upper chart it is mapped to a family of tangent circles.

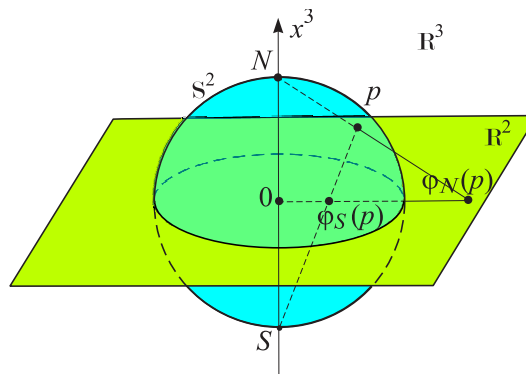


Figure 4. Stereographic projection of the sphere.

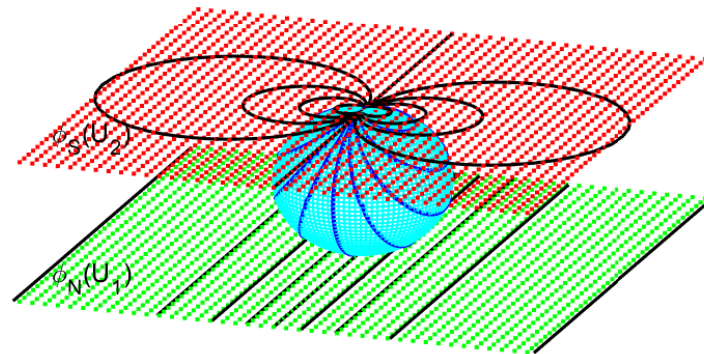


Figure 5. An atlas of the sphere.

The definition of a topological space does not allow one to define differentiable functions or other concepts from mathematical analysis on a manifold. However, many important applications of manifolds involve mathematical analysis. For example, the application of manifold theory in geometry includes properties such as volume and curvature. Typically, volumes are computed by integration, while curvatures are determined through differentiation, so extending these concepts to manifolds requires a way to make integration and differentiation meaningful on a manifold.

Similarly, applications in classical mechanics involve solving ordinary differential equations on manifolds. To give these concepts meaning, it is necessary to define an additional structure on the manifold. In order to make sense of derivatives of real-valued functions, curves, or manifolds, it is necessary to introduce a new type of manifold called a *smooth manifold*. This is a topological manifold equipped with an additional structure compatible with its topology, which allows one to determine which functions to or from the manifold are smooth.

Let \mathcal{M} be an n -dimensional topological manifold, and let \mathcal{A} be an atlas of the manifold \mathcal{M} . Consider any two charts in the atlas \mathcal{A} , denoted by (U, φ) and (V, ψ) .

Definition 3. *The coordinate transformation*

$$\psi \circ \varphi^{-1}: \varphi(U \cap V) \in \mathbb{R}^n \rightarrow \psi(U \cap V) \in \mathbb{R}^n,$$

which maps the intersection $\varphi(U \cap V)$ in \mathbb{R}^n to $\psi(U \cap V)$ in \mathbb{R}^n , is called **smooth** if the transition functions

$$\begin{aligned} y^1 &= y^1(x^1, \dots, x^n), \\ &\vdots \\ y^n &= y^n(x^1, \dots, x^n) \end{aligned}$$

have continuous partial derivatives of all orders (that is, they are infinitely differentiable, belonging to the class C^∞ in the open set $\varphi(U \cap V)$), and the determinant

$$\det \left(\left[\frac{\partial y_i}{\partial x_j} \right] \right),$$

of the Jacobian matrix of the transformation is nonzero.

This also implies that the inverse coordinate transformation of $(\psi \circ \varphi^{-1})^{-1} = \varphi \circ \psi^{-1}$ is smooth, since the corresponding transition functions

$$\begin{aligned} x^1 &= x^1(y^1, \dots, y^n), \\ &\vdots \\ x^n &= x^n(y^1, \dots, y^n) \end{aligned}$$

have continuous partial derivatives of all orders in the open set $\psi(U \cap V)$.

If the coordinate transformation is smooth, we say that the corresponding charts are smoothly compatible.

An atlas \mathcal{A} is called **smooth** if:

1. Any two charts in the atlas are smoothly compatible.
2. Every point $p \in \mathcal{M}$ lies in the domain of at least one chart.

A smooth atlas \mathcal{A} on \mathcal{M} is called **maximal** if it is not properly contained in any larger smooth atlas. This means that any chart that is smoothly compatible with every chart in \mathcal{A} is already included in \mathcal{A} . A maximal atlas is also called a **complete atlas**.

Definition 4. An n -dimensional topological manifold \mathcal{M} is said to have a **smooth structure** if there exists an atlas \mathcal{A} on the manifold satisfying the following properties:

1. For any two charts in \mathcal{A} , the corresponding coordinate transformation is smooth.
2. The atlas \mathcal{A} is maximal.

Definition 5. n -dimensional topological manifold that has a smooth structure is called an **n -dimensional smooth manifold**.

Example 2. (a) **(Normed vector spaces).** Let V be a finite-dimensional real vector space. Any norm on V defines a topology that is independent of the choice of norm. With this topology, V is an n -dimensional topological manifold with a natural smooth structure defined as follows. Any (ordered) basis (e^1, \dots, e^n) of V defines a basic isomorphism $E: \mathbb{R}^n \rightarrow V$ by

$$E(x) = \sum_{i=1}^n x^i e^i.$$

This map is a homeomorphism, so (V, E^{-1}) is a chart. The collection of all such charts defines a smooth structure, called the **standard smooth structure on V** .

(b) **(Matrix spaces).** Let $\mathcal{M}(m \times n, \mathbb{R})$ denote the set of $m \times n$ matrices with real entries. Since it is a real vector space of dimension $m \cdot n$ under matrix addition and scalar multiplication, $\mathcal{M}(m \times n, \mathbb{R})$ is a smooth $m \cdot n$ -dimensional manifold. (Since the spaces $\mathcal{M}(m \times n, \mathbb{R})$ and $\mathbb{R}^{m \cdot n}$ are isometric, they can be identified by "stacking" all matrix entries into a single row or column.) A chart on this manifold is given by

$$\varphi: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \cdot n}: X \mapsto \text{vec}(X),$$

where $\text{vec}(X)$ denotes the vector obtained from the matrix elements as described above. The dimension of this manifold is $m \cdot n$. The space $\mathbb{R}^{m \times n}$ can be equipped with a Euclidean structure via the inner product

$$\langle M_1, M_2 \rangle = \text{vec}(M_1)^T \text{vec}(M_2) = \text{tr}(M_1^T M_2).$$

The norm induced by this inner product is the Frobenius norm, defined by

$$\|M\|_F^2 = \text{tr}(M^T M),$$

i.e., $\|M\|_F^2$ is the sum of the squares of all entries of M .

Similarly, the space $\mathcal{M}(m \times n, \mathbb{C})$ of $m \times n$ complex matrices is a vector space of dimension $2mn$ over \mathbb{R} and therefore is a smooth manifold of dimension $2 \cdot m \cdot n$. In the special case $m = n$ (square matrices), the notations $\mathcal{M}(m \times n, \mathbb{R})$ and $\mathcal{M}(m \times n, \mathbb{C})$ are abbreviated as $\mathcal{M}(n, \mathbb{R})$ and $\mathcal{M}(n, \mathbb{C})$, respectively.

- (c) **(General linear group).** The **general linear group** $GL(n, \mathbb{R})$ is the set of invertible $n \times n$ matrices with real entries. It is an n^2 -dimensional manifold because it is an open subset of the n^2 -dimensional vector space $\mathcal{M}(n, \mathbb{R})$, namely the set where the (continuous) determinant function $\det(A)$ is nonzero.
- (d) **(Spaces of linear maps).** Let V and W be finite-dimensional real vector spaces, and let $L(V, W)$ denote the set of linear maps from V to W . Since $L(V, W)$ is itself a finite-dimensional vector space (whose dimension is the product of the dimensions of V and W), it naturally carries the structure of a smooth manifold, just as in item (b).

A **submanifold** of a manifold \mathcal{M} is a subset $\mathcal{P} \subset \mathcal{M}$ that itself has the structure of a manifold. The sphere \mathbb{S}^2 , defined by the equation $x^2 + y^2 + z^2 = 1$, is an example of a subset of the coordinate space \mathbb{R}^3 that inherits the natural manifold topology from \mathbb{R}^3 .

In a number of important applications of manifolds, one encounters spaces that would be smooth manifolds except that they have a “boundary” of some kind. Elementary examples of such spaces are closed intervals in \mathbb{R} , closed balls in \mathbb{R}^n , and closed hemispheres in \mathbb{S}^n . The study of **manifolds with boundary** requires a generalization of the definition of a manifold; see, for example, [6], Ch. 1.

An overview of the historical development of the concept of a manifold can be found in [16].

2.2. Smooth Maps

Let $U \subset \mathbb{R}^n$ be an open set. Any mapping $f : U \rightarrow \mathbb{R}^m$ can be represented as an ordered collection of m functions:

$$f(x^1, \dots, x^n) = (f^1(x^1, \dots, x^n), f^2(x^1, \dots, x^n), \dots, f^m(x^1, \dots, x^n)).$$

The mapping $f : U \rightarrow \mathbb{R}^m$ is called **smooth** ($f \in C^\infty$), if each function f^k , $k = 1, 2, \dots, m$, has continuous partial derivatives of all orders. Mappings with only continuous functions are called C^0 mappings. In the case where all functions f^k are analytic (a function is called **analytic** if its Taylor series converges to it in a neighborhood of each point), the mapping f is called **analytic** ($f \in C^\omega$). We have the inclusion $C^\infty \supset C^\omega$.

Let \mathcal{M} and \mathcal{P} be smooth manifolds of dimensions n and k , respectively, and let $f : \mathcal{M} \rightarrow \mathcal{P}$ be an arbitrary map. The map f is called **smooth at point** p of the manifold \mathcal{M} if there exist a local chart (U, φ) containing p and a local chart (V, ψ) containing $f(p)$ such that $f(U) \subseteq V$ and the composition

$$\psi \circ f \circ \varphi^{-1}$$

is a smooth map from $\varphi(U)$ to $\psi(V)$ (Figure 6).

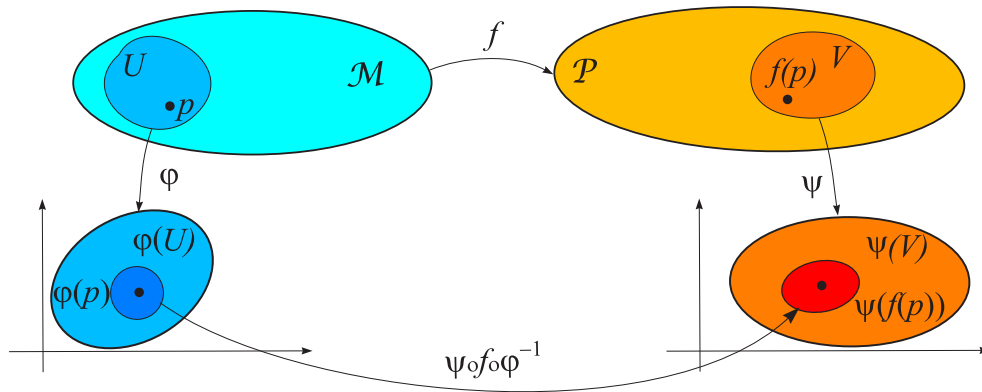


Figure 6. Smooth map between two manifolds.

A map f is called *smooth* if it is smooth at every point p of the manifold \mathcal{M} (see Figure 7).

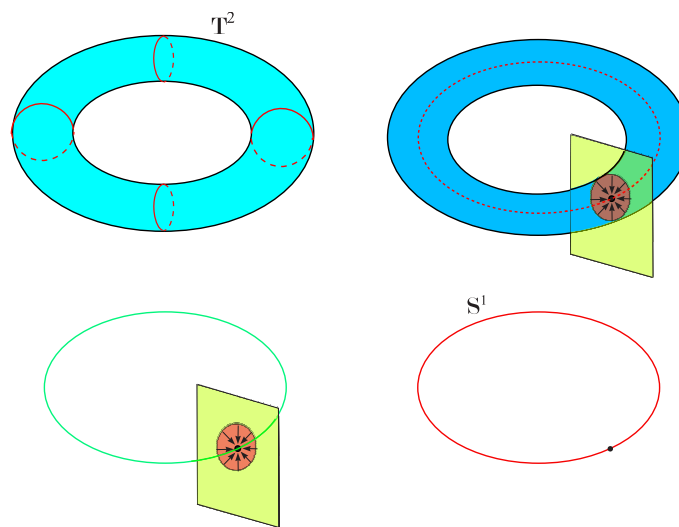


Figure 7. The projection of the torus \mathbb{T}^2 onto the circle \mathbb{S}^1 is a smooth map.

Every smooth map is continuous.

The set of smooth maps of the form $f: \mathcal{M} \rightarrow \mathcal{P}$ is denoted by $C^\infty(\mathcal{M}, \mathcal{P})$.

Let \mathcal{M} and \mathcal{P} be smooth manifolds.

A smooth map $f: \mathcal{M} \rightarrow \mathcal{P}$ is called a *diffeomorphism* if it is bijective and its inverse $f^{-1}: \mathcal{P} \rightarrow \mathcal{M}$ is also smooth.

The manifold \mathcal{M} is said to be *diffeomorphic* to the manifold \mathcal{P} if there exists a diffeomorphism $f: \mathcal{M} \rightarrow \mathcal{P}$. This is denoted by $\mathcal{M} \cong \mathcal{P}$.

Example 3. In Figure 8, an open disk and an open ellipse are shown, which are diffeomorphic. The disk is mapped to the ellipse by the smooth map

$$f: (x, y) \mapsto (x/a, y/b),$$

and the ellipse is mapped back to the disk by the inverse map

$$f^{-1}: (x, y) \mapsto (ax, by),$$

which is also smooth.

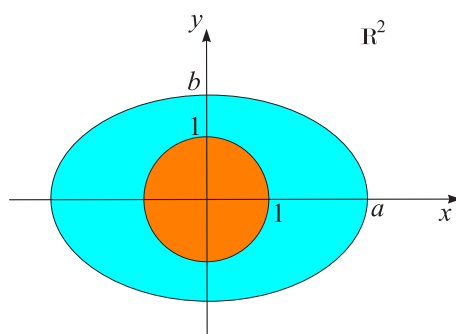


Figure 8. The ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} < 1$ and the open disk $x^2 + y^2 < 1$ are diffeomorphic.

Similar to the case when two topological spaces are considered “the same” if they are homeomorphic, two smooth manifolds are regarded as indistinguishable if they are diffeomorphic. A central question in the theory of smooth manifolds is the study of properties of smooth manifolds that are preserved under diffeomorphisms.

2.3. Tangent Space

In the study of the metric properties of regions in Euclidean space, an important role is played by properties that are defined in a theoretically infinitesimal neighborhood of a fixed point, by neglecting quantities of higher order relative to the distance to that point. Similarly, in the study of smooth manifolds it is appropriate to neglect infinitesimal quantities of higher order in order to simplify the analysis of a given problem. One way to achieve this is to introduce special concepts analogous to tangent vectors to curves and tangent planes to surfaces, as used in mathematical analysis.

Let v_p be a vector in the coordinate space \mathbb{R}^n based at a point $p \in \mathbb{R}^n$. Then, for every smooth function f defined in a neighborhood of the point p , the **directional derivative** determined by the vector v_p is defined as follows:

$$\frac{\partial f}{\partial v} \Big|_p = \lim_{t \rightarrow 0} \frac{f(p + tv_p) - f(p)}{t} = \frac{d}{dt} f(p + tv_p) \Big|_{t=0},$$

where $t \geq 0$ is a numerical parameter (in analysis, one usually considers a vector v_p of unit length). In a coordinate system, we have the formula:

$$\frac{\partial f}{\partial v} \Big|_p = \frac{\partial f}{\partial x^1} \Big|_p v^1 + \dots + \frac{\partial f}{\partial x^n} \Big|_p v^n = \langle \text{grad}(f(p)), v \rangle_p,$$

where

$$\text{grad}(f) = \left(\frac{\partial f}{\partial x^1}, \dots, \frac{\partial f}{\partial x^n} \right)$$

is the **gradient** of the function f , and $x^1, \dots, x^n, v^1, \dots, v^n$ are the coordinates of the point p and the vector v_p , respectively. Furthermore, the quantity

$$\frac{\partial f}{\partial v} \Big|_p = \sum_{i=1}^n v^i \frac{\partial}{\partial x^i} f(p)$$

will be called the **derivative of the smooth function f in the direction of the vector v_p at the point p** and will be denoted by $D_p^v f$. In this notation, it is understood that the partial derivatives are evaluated at p , since v_p is a vector at p . Note that $D_p^v f$ is a number, not a function. We write

$$D_p^v = \sum_{i=1}^n v^i \frac{\partial}{\partial x^i} \Big|_p$$

for the map that sends the function f to the number $D_p^v f$.

In this way, given a vector v_p at the point p , an operation is defined on the set $C^\infty(p)$ of smooth functions in a neighborhood of p :

$$D_p^v: C^\infty(p) \rightarrow \mathbb{R},$$

performed according to the rule

$$f(p) \mapsto D_p^v f.$$

Definition 6. A *tangent vector* at point p of the manifold \mathcal{M} is a rule

$$v_p: C^\infty(p) \rightarrow \mathbb{R},$$

which assigns to each function f in the set $C^\infty(p)$ a number $v_p f$ satisfying the following properties:

1. $v_p(f + g) = v_p f + v_p g$,
2. $v_p(\alpha f) = \alpha v_p f$,
3. $v_p(f \cdot g) = f(p) \cdot v_p g + v_p f \cdot g(p)$,

where $f, g \in C^\infty(p)$, $\alpha \in \mathbb{R}$.

The association $v_p \mapsto D_p^v$ of the directional derivative D_p^v with the tangent vector v_p allows tangent vectors to be characterized as certain operators acting on functions.

The set of vectors that are tangent to the manifold \mathcal{M} at a point p is denoted by $T_p \mathcal{M}$. We now define on this set the operations of addition of tangent vectors and scalar multiplication of a tangent vector.

Let $v_p, w_p \in T_p \mathcal{M}$, $f \in C^\infty(p)$ and $\alpha \in \mathbb{R}$. We define

$$\begin{aligned} (v_p + w_p)f &= v_p f + w_p f, \\ (\alpha v_p)f &= \alpha v_p f. \end{aligned}$$

It is not difficult to verify that both the *sum* of the tangent vectors v_p and w_p , and the *product* of a tangent vector with a scalar αv_p , are also tangent vectors. In this way, the set $T_p \mathcal{M}$ becomes a vector space. It is called the *tangent space* to the smooth manifold \mathcal{M} at the point p .

Let (U, φ) be a local chart (coordinate system) and let $p \in U$. For every function $f \in C^\infty(p)$ one can construct a smooth function $\hat{f} = f \circ \varphi^{-1}$, defined on the open subset $\varphi(U)$ of \mathbb{R}^n (see Figure 9).

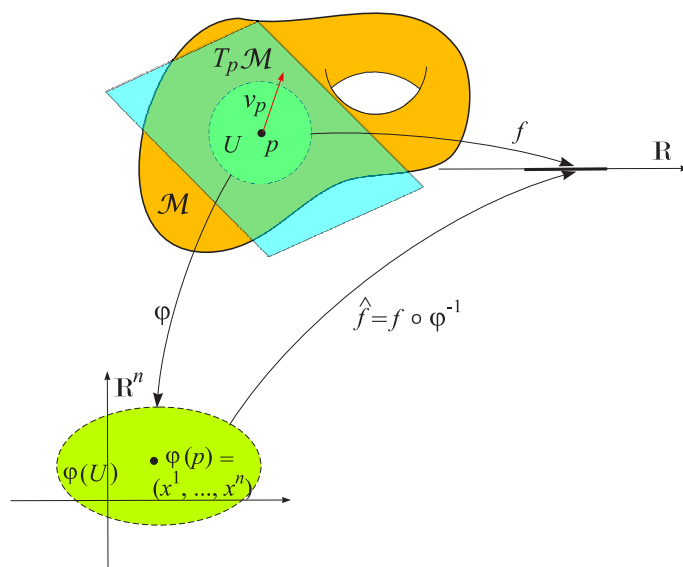


Figure 9. Tangent space of a manifold.

By computing the partial derivatives of \hat{f} with respect to the variables x^1, \dots, x^n , we obtain

$$\frac{\partial}{\partial x^i} \hat{f}(\varphi(p)) = \frac{\partial}{\partial x^i} \Big|_{\varphi(p)} (f \circ \varphi^{-1}), \quad i = 1, \dots, n.$$

In this way, for each function $f \in C^\infty(p)$, one can associate n numbers

$$f \mapsto \frac{\partial}{\partial x^1} \Big|_{\varphi(p)} (f \circ \varphi^{-1}), \dots, \frac{\partial}{\partial x^n} \Big|_{\varphi(p)} (f \circ \varphi^{-1}).$$

We see that the choice of a system of local coordinates determines n vectors in the tangent space $T_{\varphi(p)}\mathcal{M}$ acting according to the rule

$$f \mapsto \frac{\partial}{\partial x^i} \Big|_{\varphi(p)} (f \circ \varphi^{-1}), \quad i = 1, \dots, n.$$

These vectors are conventionally denoted by

$$\frac{\partial}{\partial x^i} \Big|_p.$$

The quantities $\partial/\partial x^i|_p$ represent linear partial differential operators, which act according to the rule

$$\frac{\partial}{\partial x^i} \Big|_p f = \frac{\partial f}{\partial x^i} \Big|_p = \frac{\partial}{\partial x^i} \Big|_{\varphi(p)} (f \circ \varphi^{-1}) = \frac{\partial \hat{f}}{\partial x^i}(\hat{p}).$$

Theorem 1. [6], Ch. 3 *The vectors*

$$\frac{\partial}{\partial x^1} \Big|_p, \dots, \frac{\partial}{\partial x^n} \Big|_p$$

form a basis of the tangent space $T_p\mathcal{M}$.

In this way, we obtain the important result that the dimension of the tangent space $T_p\mathcal{M}$ is equal to the dimension of the manifold \mathcal{M} ,

$$\dim(T_p\mathcal{M}) = \dim(\mathcal{M}).$$

The orthogonal complement of the tangent space is called the *normal space* and is denoted by $N_p\mathcal{M}$.

Instead of working with the dimension of the manifold \mathcal{M} , it is often more convenient to use the *codimension* of \mathcal{M} , denoted by $\text{codim}(\mathcal{M})$, which is equal to the dimension of the normal space $N_p\mathcal{M}$. Since

$$T_p\mathcal{M} + N_p\mathcal{M} = \mathcal{S},$$

where \mathcal{S} is the ambient manifold, we have

$$\dim(\mathcal{M}) + \text{codim}(\mathcal{M}) = \dim(\mathcal{S}).$$

2.4. Differential of a Map

To analyze the action of smooth maps on tangent vectors, it is necessary to consider the differentiation of such maps. In the case of a smooth manifold between Euclidean spaces, the total derivative of the map at a point (represented by its Jacobian matrix) is a linear map that provides the "best linear approximation" of the map near the given point. In the case of manifolds, a linear map is defined between the tangent spaces.

Definition 7. Let \mathcal{M} and \mathcal{P} be smooth manifolds, and let $f: \mathcal{M} \rightarrow \mathcal{P}$ be a smooth map. The **differential** of f at a point $p \in \mathcal{M}$ is the linear map from the tangent space of \mathcal{M} to the tangent space of \mathcal{P} ,

$$D_p f: T_p \mathcal{M} \rightarrow T_{f(p)} \mathcal{P},$$

defined as follows (see Figure 10). Let $v \in T_p \mathcal{M}$. Consider a curve $\varphi: \mathbb{R} \rightarrow \mathcal{M}$ with $\varphi(0) = p$ whose tangent vector is v . Then $D_p f v$ is the tangent vector to the curve $f \circ \varphi: \mathbb{R} \rightarrow \mathcal{P}$.

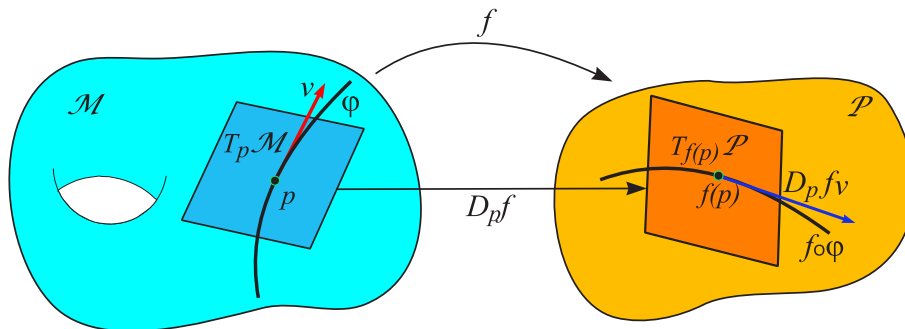


Figure 10. Differential of a smooth map.

Proposition 1. (Properties of the differential). Let \mathcal{M} , \mathcal{P} , and \mathcal{Z} be smooth manifolds. Let $f: \mathcal{M} \rightarrow \mathcal{P}$ and $g: \mathcal{P} \rightarrow \mathcal{Z}$ be smooth maps, and let $p \in \mathcal{M}$.

- $D_p f: T_p \mathcal{M} \rightarrow T_{f(p)} \mathcal{P}$ is a linear map.
- $D_p (g \circ f) = D_{f(p)} g \circ D_p f: T_p \mathcal{M} \rightarrow T_{g \circ f(p)} \mathcal{Z}$.
- $D_p (i_{\mathcal{M}}) = i_{T_p \mathcal{M}}: T_p \mathcal{M} \rightarrow T_p \mathcal{M}$, where $i_{\mathcal{M}}$ is the identity map on \mathcal{M} .
- If f is a diffeomorphism, then $D_p f: T_p \mathcal{M} \rightarrow T_{f(p)} \mathcal{P}$ is an isomorphism, and $(D_p f)^{-1} = D_{f(p)} (f^{-1})$.

Using (x^1, \dots, x^n) to denote coordinates in the domain of f and (y_1, \dots, y^m) to denote coordinates in the codomain, the action of $D_p f$ on a typical basis vector is

$$D_p f \left(\frac{\partial}{\partial x^i} \Big|_p \right) = \frac{\partial f^j}{\partial x^i} (p) \frac{\partial}{\partial y^j} \Big|_{f(p)}. \quad (1)$$

Therefore, the matrix of $D_p f$ in terms of the coordinate bases is

$$\begin{bmatrix} \frac{\partial f_1}{\partial x^1} (p) & \cdots & \frac{\partial f_1}{\partial x^n} (p) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x^1} (p) & \cdots & \frac{\partial f_m}{\partial x^n} (p) \end{bmatrix}.$$

This matrix is precisely the **Jacobian matrix** of f at p , which is the matrix representation of the total derivative $Df(p): \mathbb{R}^n \rightarrow \mathbb{R}^m$.

Example 4. We consider the maps $f: \mathbb{B}_{\mathbb{R}}^3 \rightarrow \mathbb{R}^3$ and $g: \mathbb{R}^3 \rightarrow \mathbb{B}_{\mathbb{R}}^3$, defined by

$$f(x) = \left(\frac{x^1}{\sqrt{1-x^{12}}}, \frac{x^2}{\sqrt{1-x^{22}}}, \frac{x^3}{\sqrt{1-x^{32}}} \right),$$

$$g(x) = \left(\frac{x^1}{\sqrt{1+x^{12}}}, \frac{x^2}{\sqrt{1+x^{22}}}, \frac{x^3}{\sqrt{1+x^{32}}} \right),$$

where \mathbb{B}_R^3 is the open 3-dimensional ball of radius R . These maps are smooth, and it can be shown that they are inverses of each other for $\|x\|_2 = \sqrt{x^1{}^2 + x^2{}^2 + x^3{}^2} < 1$. Therefore, both maps are diffeomorphisms.

Let $p = (p^1, p^2, p^3)$ be a point on the sphere \mathbb{S}_R^2 that does not coincide with the north pole. To obtain the tangent space $T_{f(p)}f(\mathbb{S}_R^2)$, we need to compute the differential $D_p f: \mathbb{R}^2 \rightarrow T_{f(p)}f(\mathbb{S}_R^2)$. For a local parametrization of \mathbb{S}_R^2 and $f(\mathbb{S}_R^2)$ in \mathbb{R}^2 , we use the stereographic projections (see Example 1)

$$\begin{aligned}\varphi: \mathbb{S}_R^2 &\rightarrow \mathbb{R}^2, (x^1, x^2) = \varphi(p) = \left(\frac{p^1}{1-p^3}, \frac{p^2}{1-p^3} \right), \\ \varphi^{-1}: \mathbb{R}^2 &\rightarrow \mathbb{S}_R^2, (p^1, p^2, p^3) = \varphi^{-1}(x) = \frac{1}{1+x^1{}^2+x^2{}^2} (2x^1, 2x^2, x^1{}^2+x^2{}^2-1), \\ \psi: f(\mathbb{S}_R^2) &\rightarrow \mathbb{R}^2, \psi(y) = \left(\frac{y^1}{1-y^3}, \frac{y^2}{1-y^3} \right),\end{aligned}$$

where

$$y = f(p) = \left(\frac{p^1}{\sqrt{1-p^1{}^2}}, \frac{p^2}{\sqrt{1-p^2{}^2}}, \frac{p^3}{\sqrt{1-p^3{}^2}} \right).$$

The matrices of partial derivatives are

$$\begin{aligned}D_y \psi &= \begin{bmatrix} 1/(1-y^3) & 0 & -y^1/(1-y^3) \\ 0 & 1/(1-y^3) & -y^2/(1-y^3) \end{bmatrix}, \\ D_p f &= \begin{bmatrix} 1/\sqrt{1-p^1{}^2} & 0 & 0 \\ -2p^1{}^2/(1-p^1{}^2) & 1/\sqrt{1-p^2{}^2} & 0 \\ 0 & -2p^2{}^2/(1-p^2{}^2) & 1/\sqrt{1-p^3{}^2} \\ 0 & 0 & -2p^3{}^2/(1-p^3{}^2) \end{bmatrix}, \\ D_x \varphi^{-1} &= \frac{2}{(1+x^1{}^2+x^2{}^2)^2} \begin{bmatrix} 1-x^1{}^2+x^2{}^2 & -2x^1x^2 \\ -2x^1x^2 & 1+x^1{}^2-x^2{}^2 \\ 2x^1 & 2x^2 \end{bmatrix}.\end{aligned}$$

The tangent space $T_{f(p)}f(\mathbb{S}_R^2)$ is obtained as the image of \mathbb{R}^2 under the differential $D_x f = D_p f D_x \varphi^{-1}$.

Figure 11 shows the diffeomorphic image $\mathbb{S}_R^2 \mapsto f(\mathbb{S}_R^2)$ of a sphere with radius $R = 0.97$. The tangent spaces $T_p \mathbb{S}_R^2$ and $T_{f(p)}f(\mathbb{S}_R^2)$ are computed for the points $p = R(1/\sqrt{3}, -1/\sqrt{3}, 1/\sqrt{3})$ and $x = \varphi(p)$, $y = f(p)$.

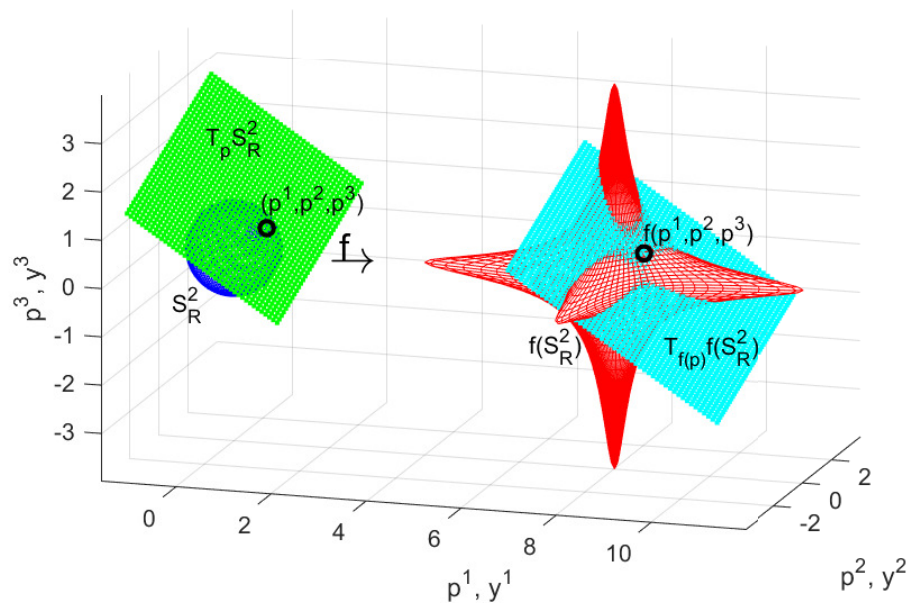


Figure 11. Tangent space of a diffeomorphic map of a sphere.

The matrix of the differential $D_p \hat{f}: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is obtained from

$$D_p \hat{f} = D_y \psi D_p f D_x \varphi^{-1}.$$

Very good descriptions of tangent vectors and tangent spaces of manifolds, accompanied by examples, are given in [6], Ch. 3, [10], Ch. 1, [17], [9], Ch. 3, [12], Ch. 4, [18], Ch. 4.

2.5. Tangent and Normal Bundle

Let \mathcal{M} be a smooth manifold. Let us define the disjoint union of all tangent spaces of \mathcal{M} ,

$$T\mathcal{M} = \bigcup_{p \in \mathcal{M}} T_p \mathcal{M}.$$

The set $T\mathcal{M}$ is called **tangent bundle** of the manifold \mathcal{M} . The term **bundle** indicates that $T\mathcal{M}$ consists of “layers”—the tangent spaces at the individual points of the manifold. The tangent bundle of an m -dimensional manifold in \mathbb{R}^n is itself a manifold whose dimension is equal to $2m$.

In the trivial case, when $\mathcal{M} = \mathbb{R}^n$, for each $p \in \mathbb{R}^n$ the tangent space $T_p \mathbb{R}^n$ can be identified with \mathbb{R}^n . Therefore, in this case we have $T\mathbb{R}^n \cong \mathbb{R}^n \times \mathbb{R}^n$, i.e., the tangent bundle $T\mathbb{R}^n$ is diffeomorphic to the Cartesian product $\mathbb{R}^n \times \mathbb{R}^n$.

The only tangent bundles that can be easily visualized are those of the real line \mathbb{R} and of the circle. The tangent bundle of two-dimensional manifolds is four-dimensional and therefore difficult to visualize.

The tangent bundle of the real line coincides with \mathbb{R}^2 . The tangent bundle of the circle is obtained by considering all tangent spaces (Figure 12, top) and combining them disjointly into a smooth manifold (Figure 12, bottom).

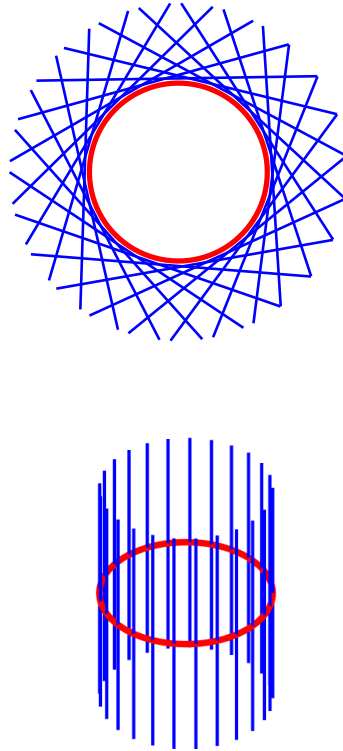


Figure 12. The tangent bundle of a circle.

The map $\pi: T\mathcal{M} \rightarrow \mathcal{M}$, which assigns to each tangent vector v the point $p \in U$, at which the vector is tangent to \mathcal{M} ($v \in T\mathcal{M}$), is called the *the natural projection*. The preimage of a point $p \in \mathcal{M}$ under the natural projection, $\pi^{-1}(p)$, is the tangent space $T_p\mathcal{M}$. This space is called the *fiber of the bundle over the point p* .

Let \mathcal{M} be a smooth manifold. Let us define the disjoint union of all normal spaces of \mathcal{M} ,

$$N\mathcal{M} = \bigcup_{p \in \mathcal{M}} N_p\mathcal{M}.$$

The set $N\mathcal{M}$ is called *normal bundle*. The normal bundle of a circle is shown in Figure 13.

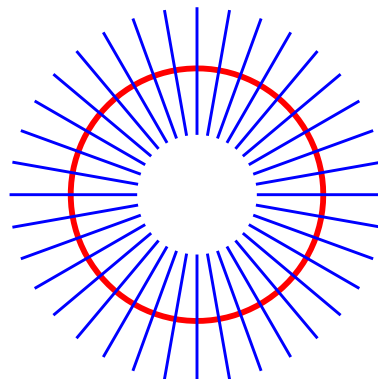


Figure 13. The normal bundle of a circle.

Assume that $\mathcal{M} \subseteq \mathbb{R}^n$ is an m -dimensional submanifold. The normal bundle $N\mathcal{M}$ consists of all vectors that are normal to \mathcal{M} :

$$N\mathcal{M} = \{(p, v) \in \mathbb{R}^n \times \mathbb{R}^n : p \in \mathcal{M}, v \in N_p\mathcal{M}\}.$$

In this case, $N\mathcal{M}$ can be viewed as an n -dimensional submanifold of the Cartesian product $\mathbb{R}^n \times \mathbb{R}^n$.

Let $\mathcal{M} \subseteq \mathbb{R}^n$ be an m -dimensional submanifold without boundary. Then $N\mathcal{M}$ is a smooth manifold of dimension n .

Normal bundles can be considered more generally when we have a submanifold $\mathcal{P} \subset \mathcal{M}$, in order to understand the geometry of \mathcal{P} within \mathcal{P} in \mathcal{M} .

Definition 8. (Normal bundle of a submanifold). Let $\mathcal{M} \subset \mathbb{R}^m$ be a manifold without boundary, and let \mathcal{P} be a submanifold of \mathcal{M} . The normal bundle of \mathcal{P} in \mathcal{M} is defined as the set

$$N(\mathcal{P}, \mathcal{M}) := \{(p, v) : p \in \mathcal{P}, v \in T_p\mathcal{M} \text{ and } v \perp T_p\mathcal{P}\}.$$

The normal bundle $N(\mathcal{P}, \mathcal{M})$ is a smooth manifold of dimension equal to $\dim(\mathcal{M})$.

2.6. Tubular Neighborhoods

In this section we consider an important application of normal bundles, which is characterized by the fact that every smooth manifold \mathcal{M} without boundary possesses a special type of neighborhood.

Generally speaking, a **tubular neighborhood** U of a smooth submanifold \mathcal{M} is an open set around the submanifold whose structure resembles that of the normal bundle. This definition can be made more concrete by the following example. Let us consider a smooth plane curve without self-intersections. At each point of the curve we draw a straight line perpendicular to the curve. Except in the case where the manifold is a straight line, these lines will intersect in a complicated way (Figure 14). However, if we consider a narrow strip around the curve, the portions of the normal lines contained in this strip will not intersect and will cover the entire strip without gaps.

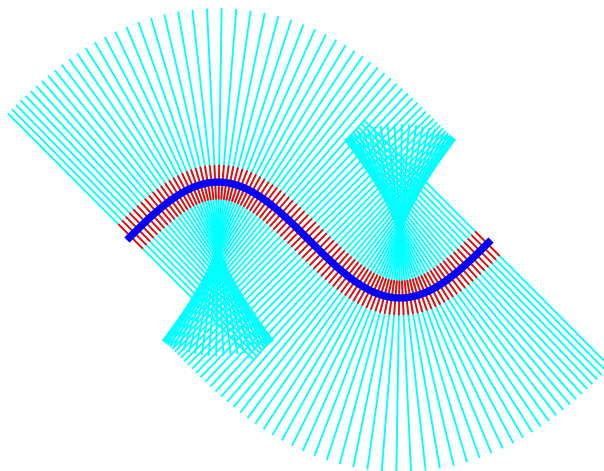


Figure 14. A tubular neighborhood in \mathbb{R}^2 with normal lines.

The tubular neighborhood of a space curve in \mathbb{R}^3 is shown in Figure 15.

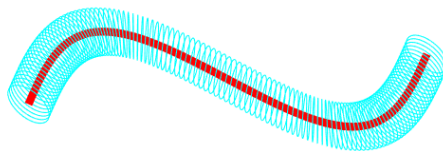


Figure 15. A tubular neighborhood in \mathbb{R}^3 .

Let $\mathcal{M} \subseteq \mathbb{R}^n$ be an m -dimensional submanifold. Viewing the normal bundle $N\mathcal{M}$ as a submanifold of $\mathbb{R}^n \times \mathbb{R}^n$, we define the smooth map

$$E: N\mathcal{M} \rightarrow \mathbb{R}^n, (p, v) \mapsto p + v.$$

It maps each normal space $N_p\mathcal{M}$ affinely through p and orthogonally to $T_p\mathcal{M}$. The tubular neighborhood of \mathcal{M} is the neighborhood U of \mathcal{M} in \mathbb{R}^n which is the diffeomorphic image, under this map, of an open subset $V \subset N\mathcal{M}$ of the form

$$V = \{(p, v) \in N\mathcal{M} : |v| < \varepsilon(p)\},$$

for some positive continuous function $\varepsilon: \mathcal{M} \rightarrow \mathbb{R}^+$ (Figure 16). We have the following definition:

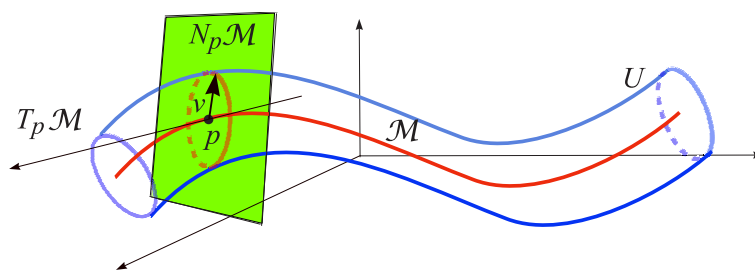


Figure 16. A tubular neighborhood.

Definition 9. (Tubular neighborhood). Let $\mathcal{M} \subset \mathbb{R}^n$ be a smooth manifold without boundary. A **tubular neighborhood** of \mathcal{M} is an open subset U of \mathbb{R}^n containing \mathcal{M} such that E maps an open subspace $V \subset N\mathcal{M}$ diffeomorphically onto U , where V is defined by a smooth function $\varepsilon: \mathcal{M} \rightarrow \mathbb{R}^+$.

The key property of smooth manifolds embedded in a Euclidean space is that they always possess a tubular neighborhood.

Theorem 2. Every embedded submanifold of \mathbb{R}^n has a tubular neighborhood.

Tubular neighborhoods and their properties are studied in detail in [6], Ch. 6, [10], Ch. 6, [13], Ch. 5.

2.7. Singular and Regular Points

Definition 10. Let f be a smooth map from an m -dimensional manifold \mathcal{M} to an n -dimensional manifold \mathcal{P} .

- (a) A point $p \in \mathcal{M}$ is called **critical** or **singular point** of f if the derivative $D_p f: T_p \mathcal{M} \rightarrow T_{f(p)} \mathcal{P}$ is not surjective; that is, if the rank of the Jacobian matrix $\left(\frac{\partial f}{\partial x}\right)\Big|_p$ is smaller than the dimension n of \mathcal{P} . The image $f(p)$ of a critical point is called a **critical value** of f .
- (b) A point $p \in \mathcal{M}$ is called a **regular point** of f if it is not critical. A point $q \in \mathcal{P}$ is called a **regular value** of f if its inverse image $f^{-1}(q)$ contains no critical points.

Note that if the dimension of \mathcal{M} is smaller than the dimension of \mathcal{P} , then all points of \mathcal{M} are critical points of f . On the other hand, if $f(\mathcal{M})$ is not the whole of \mathcal{P} , then all points of $\mathcal{P} \setminus f(\mathcal{M})$ are regular values.

Example 5. (a) Let \mathcal{M} be a torus embedded as a submanifold in three-dimensional Euclidean space. From Figure 17 it can be seen that there exist exactly four horizontal planes (that is, planes of the form $z = \text{const}$), denoted H_1, H_2, H_3, H_4 , which are tangent planes to \mathcal{M} at the points p_1, p_2, p_3, p_4 , respectively. This corresponds to the fact that the function z restricted to \mathcal{M} has critical points at p_1, p_2, p_3, p_4 . These critical points correspond to the critical values $z_1 = f(p_1), z_2 = f(p_2), z_3 = f(p_3), z_4 = f(p_4)$.

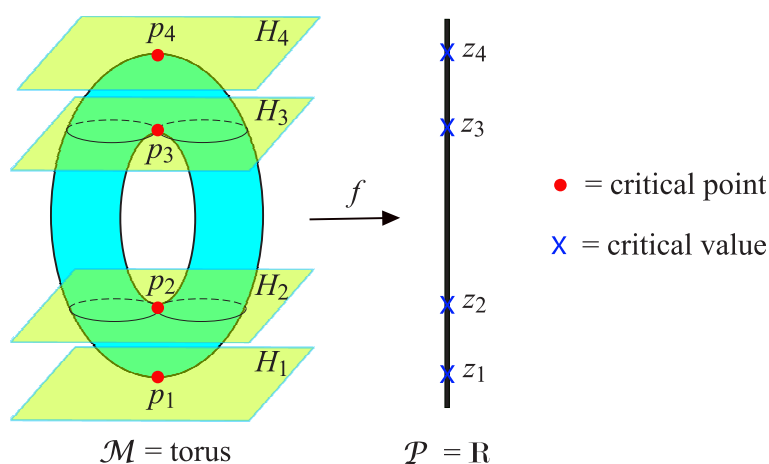


Figure 17. Critical points and values of the function $f(x, y, z) = z$.

(b) Let $f: \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ be the mapping

$$x = (x^1, \dots, x^{n+1}) \mapsto |x|^2 = x^{1^2} + \dots + x^{n+1^2}.$$

The derivative $\frac{\partial f}{\partial x}$ at point $x = (x^1, \dots, x^{n+1})$ is the linear mapping given, in the standard basis, by the matrix $(2x^1, \dots, 2x^{n+1})$. Thus, $D_x f: \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ is surjective unless $f(x) = 0$, so every nonzero real number is a regular value of f . In particular, we obtain the sphere $\mathbb{S}^n = f^{-1}(1)$ as an n -dimensional manifold.

(c) We consider the case in which the full-rank condition for the Jacobian matrix is not satisfied. Let the manifold $\mathcal{M} \subset \mathbb{R}^3$ be defined by the equation $x^{1^2} - x^{2^2} - x^{3^2} = 0$ (see Figure 18). On the set $\mathcal{M} \setminus 0$, a structure of a two-dimensional C^∞ submanifold can be defined. At the point 0, all minors of the Jacobian matrix vanish and its rank is not maximal. Therefore, the set \mathcal{M} is an algebraic variety [34], and the point 0 is a singular point of this variety.

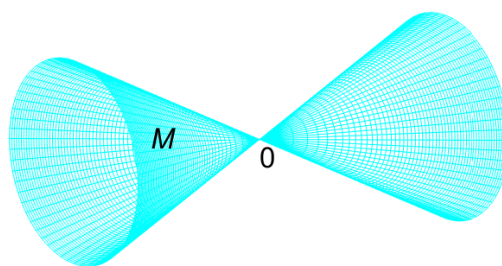


Figure 18. The case of a singular point.

The study of abrupt changes that arise in families of objects depending smoothly on parameters is the subject of *singularity theory* [19–22]. This theory deals with the classification of types of changes and the characterization of the sets of parameters that give rise to sudden transitions. Singularity theory forms the foundation of the famous *catastrophe theory* [23–28].

2.8. Sard's Theorem and Morse Functions

The following result shows that almost every point in the image of a smooth map is a regular value.

Theorem 3. (Sard's Theorem). *Suppose that \mathcal{M} and \mathcal{P} are smooth n -manifolds and $f: \mathcal{M} \rightarrow \mathcal{P}$ is a smooth map. Then the set of critical values of f has measure zero in \mathcal{P} .*

The proof of Theorem 3 can be found in ([10], Ch. 1).

Sard's Theorem is illustrated in Figure 19. This theorem is a key result in differential topology and is applied in many situations. For example, it is crucial for Thom's Transversality Theorem (see Theorem 4).

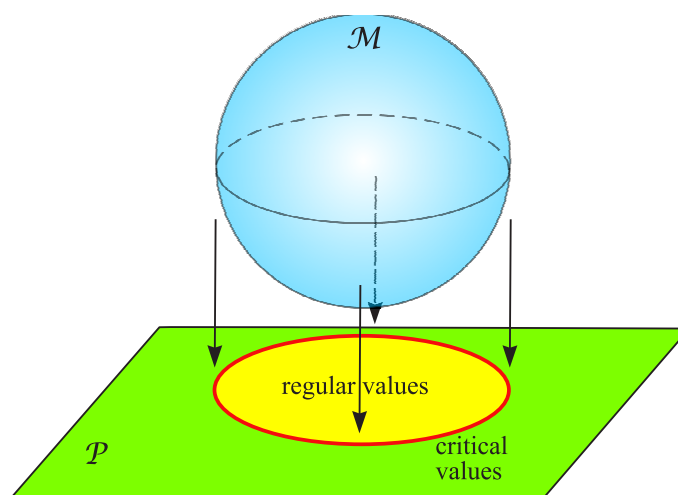


Figure 19. Sard's Theorem.

An equivalent formulation of Sard's Theorem is as follows:

If $f: \mathcal{M} \rightarrow \mathcal{P}$ is a smooth map between manifolds, then almost every $y \in \mathcal{P}$ is a regular value of f , i.e., the set of regular values is a dense subset of \mathcal{P} , or equivalently, every open subset of \mathcal{P} contains a regular value.

It should be emphasized that this theorem refers to regular values, not regular points. For example, a constant function of one variable has no regular points (all points are critical), but it has only one critical value, so the remaining points in \mathbb{R} are regular values. A set consisting of a single point clearly has measure zero.

Sard's lemma was published by the American mathematician Arthur Sard in 1942.

It is interesting to study the behavior of a function f near its critical points. If \mathcal{M} is a compact set, every function on it must have a maximum and a minimum. However, if $f(x)$ has an extremum, its derivative $D_x f$ must be equal to zero. Thus, over a compact domain, every function has at least two critical points (except in the case $\mathcal{M} = \{\text{a single point}\}$).

Let us consider a smooth function $f: \mathbb{R}^k \rightarrow \mathbb{R}$. Locally, near a point $c \in \mathbb{R}^k$, f can be expressed using the Taylor series:

$$f(x) = f(c) + \sum_{i=1}^k \frac{\partial f}{\partial x^i}(c) \cdot (x^i - c^i) + \frac{1}{2} \sum_{i,j=1}^k \frac{\partial^2 f}{\partial x^i \partial x^j}(c) \cdot (x^i - c^i)(x^j - c^j) + o(|x^3|).$$

If c is a critical point, then by definition:

$$D_c f = \left(\frac{\partial f}{\partial x^1}(c), \dots, \frac{\partial f}{\partial x^k}(c) \right) = (0, \dots, 0).$$

Therefore, in the neighborhood of a critical point, we have

$$f(x) = f(c) + \frac{1}{2} \sum_{i,j=1}^k \frac{\partial^2 f}{\partial x^i \partial x^j}(c) \cdot (x^i - c^i)(x^j - c^j) + o(|x^3|). \quad (2)$$

Hence, the best possible approximation of the local behavior of f at the point c is given by the **Hessian matrix** of second derivatives:

$$H = \left[\frac{\partial^2 f}{\partial x^i \partial x^j} \right].$$

Note that the Hessian H is a real symmetric matrix and therefore has only real eigenvalues. At a point where all eigenvalues of the Hessian are positive, the function f has a minimum; at a point where they are negative, f has a maximum.

Definition 11. (Non-degenerate critical points and Morse functions). For a smooth function $f: \mathbb{R}^k \rightarrow \mathbb{R}$, a point $c \in \mathbb{R}^k$ where $D_c f = 0$ but the Hessian matrix

$$H_c(f) = \left[\frac{\partial^2 f}{\partial x^i \partial x^j}(c) \right]$$

is invertible at c , is called **non-degenerate critical point** of f . If all critical points of f are non-degenerate, then f is called a **non-degenerate function** or **Morse function**.

Morse's lemma was published by the American mathematician Marston Morse in 1925.

Computations in the neighbourhood of algebraic singularities are considered in [29].

2.9. Transverse Intersection of Manifolds

Let \mathcal{M} and \mathcal{P} be smooth submanifolds of an ambient manifold \mathcal{S} . They are said to **intersect transversally** if for every $x \in \mathcal{M} \cup \mathcal{P}$, their tangent spaces at x satisfy

$$T_{\mathcal{M}}(x) + T_{\mathcal{P}}(x) = T_{\mathcal{S}}(x).$$

(Figure 20). That is, the directions tangent to \mathcal{M} together with the directions tangent to \mathcal{P} span all possible directions of the ambient manifold.

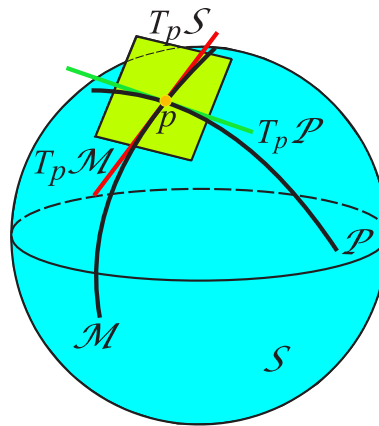


Figure 20. Transverse intersection of two manifolds.

The term frequently used synonymously for transversal is *general position*, i.e., two manifolds which intersect transversally are said to be in general position.

Two linear subspaces \mathcal{X} and \mathcal{Y} of a linear space \mathcal{L} are transverse, if their sum is equal to the whole space, $\mathcal{X} + \mathcal{Y} = \mathcal{L}$. For instance, two planes intersecting at nonzero angles in \mathbb{R}^3 are transverse (Figure 21).

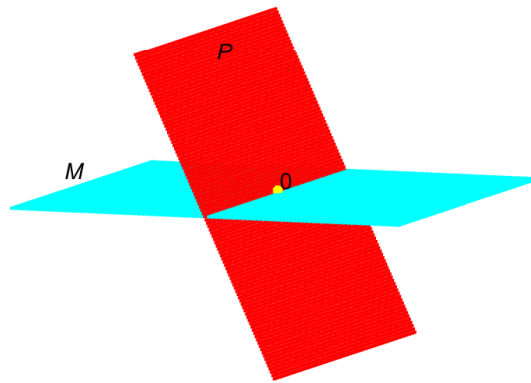


Figure 21. Transverse planes in \mathbb{R}^3 .

In three-dimensional space, *transverse curves* do not intersect. Indeed, if \mathcal{C}_1 and \mathcal{C}_2 are curves in \mathbb{R}^3 , then

$$\dim \mathcal{C}_1 + \dim \mathcal{C}_2 = 1 + 1 < 3,$$

so their tangent spaces cannot span the tangent space of the ambient manifold at a common point.

A *curve transverse to a surface* intersects the surface in isolated points. In this case

$$\dim \mathcal{C} + \dim \mathcal{S} = 1 + 2 = 3,$$

and transversality implies that at each intersection point the tangent line to the curve together with the tangent plane to the surface span \mathbb{R}^3 .

Similarly, *two surfaces transverse to each other* intersect in a curve. Indeed,

$$\dim \mathcal{S}_1 + \dim \mathcal{S}_2 = 2 + 2 > 3,$$

and their transverse intersection has dimension

$$2 + 2 - 3 = 1.$$

Note that two perpendicular planes in \mathbb{R}^3 intersect transversally, whereas two perpendicular lines lying in one and the same plane do not (Figure 22, left). Curves that are tangent to a surface at a point (for example, curves lying entirely on a surface) do not intersect the surface transversally. The same is true for planes that are tangent to a surface at a point (Figure 22, right). If an intersection of two submanifolds is transverse, then it is itself a smooth submanifold whose codimension is equal to the sums of the codimensions of the two intersecting manifolds.

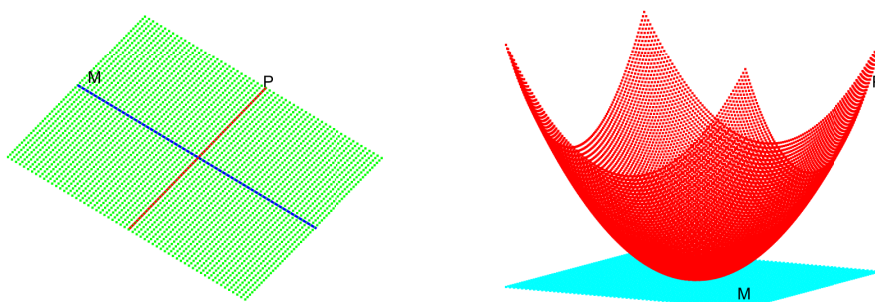


Figure 22. Intersections which are not transverse.

The following result shows that transverse intersections are generic among intersections of smooth manifolds.

Theorem 4. (Thom's Transversality Theorem). ([4], Ch. 2) *Suppose we have a family of smooth maps*

$$f_s: \mathcal{M} \rightarrow \mathcal{P}, \quad f_s(x) = F(x, s),$$

where each map depends smoothly on a parameter s belonging to a parameter space \mathcal{S} . Assume that \mathcal{M} may have a boundary, while \mathcal{P} and a given submanifold $\mathcal{Z} \subset \mathcal{P}$ do not.

If the full mapping $F(x, s)$, as well as its restriction to the boundary of \mathcal{M} , intersects \mathcal{Z} in a transversal way, then for "almost all" choices of the parameter s the corresponding map $f_s(x)$ also intersects \mathcal{Z} transversally, both in the interior of \mathcal{M} and on its boundary.

In other words, transversality is a generic property: although a particular map may fail to be transversal, a small perturbation – obtained by slightly changing the parameter s – will typically restore transversality.

The concept of transversality was developed by the French mathematician René Thom in the 1950s.

2.10. Lie Groups

Lie groups are one of the powerful tools of differential topology, applied in a variety of areas, such as the theory of differential equations, the study of special functions, and matrix analysis. In this section, some basic information about Lie groups and their properties is provided.

The theory of Lie groups and Lie algebras is presented in depth in [6], Ch. 7, [9], Ch. 4, [30], Ch. 3. A comprehensive overview of Lie group theory, matrix Lie groups, and matrix Lie algebras is given in [31]. The group-theoretic approach to matrices and vector spaces is developed in detail in [32]. Applications of Lie groups in the theory of differential equations are discussed in [33].

2.10.1. Basic Definitions

Definition 12. A **group** is a set G together with a **group operation**, usually called **multiplication**, such that for any two elements g and h in G , their product $g \cdot h$ is again an element of G . The group operation is required to satisfy the following properties:

(1) **Associativity.** If g, h , and k are elements of G , then

$$g \cdot (h \cdot k) = (g \cdot h) \cdot k.$$

(2) **Existence of an identity element.** There exists a distinguished element $e \in G$, called the **identity element**, which satisfies

$$e \cdot g = g \cdot e = g$$

for all $g \in G$.

(3) **Existence of an inverse element.** For every $g \in G$, there exists an **inverse element**, denoted g^{-1} , which satisfies

$$g \cdot g^{-1} = e = g^{-1} \cdot g.$$

Below are some elementary examples of groups.

Example 6. (a) Let $G = \mathbb{Z}$ be the set of integers with the group operation being **addition**. Clearly, associativity holds, the identity element is 0, and the “inverse” of an integer x is $-x$.

(b) Similarly, $G = \mathbb{R}$ – the set of real numbers, is also a group under addition. Again, the identity element is 0, and the inverse of a real number x is $-x$. In both cases, the group operation is commutative: $g \cdot h = h \cdot g$ for all $g, h \in G$. Such groups are called **Abelian**.

(c) Let $G = GL(n, \mathbb{Q})$ be the set of **invertible** $n \times n$ **matrices** with rational entries. The group operation is **matrix multiplication**. The identity element is the **identity matrix** I , and the inverse of a matrix A is the usual inverse matrix, whose entries are again rational numbers.

(d) Similarly, the **general linear group** $GL(n, \mathbb{R})$ – the set of invertible $n \times n$ matrices with real entries – is a group under matrix multiplication with the same identity element and inverses as in the previous example.

Lie groups are smooth manifolds that are also groups, in which the multiplication and inversion operations are smooth maps. Besides providing many interesting examples of manifolds in their own right, Lie groups are a fundamental tool in the study of more general manifolds.

In examples (b) and (d) given above, we actually have Lie groups, since the sets \mathbb{R} and $GL(n)$ are smooth manifolds. In both cases, the group operation is smooth (in fact, analytic). This leads to the following general definition of a Lie group.

Definition 13. (Lie Groups). A **Lie group** is a smooth manifold G that is also a group in the algebraic sense, with the property that the multiplication map $m : G \times G \rightarrow G$ and the inverse map $i : G \rightarrow G$, defined by

$$m(g, h) = g \cdot h, \quad i(g) = g^{-1},$$

are both smooth maps of manifolds.

In fact, one can equivalently state that if G is a smooth manifold with a group structure such that the map

$$G \times G \rightarrow G, \quad (g, h) \mapsto gh^{-1}$$

is smooth, then G is a Lie group.

Lie groups exist at the boundary between algebra and topology. The algebraic properties of Lie groups follow from the group axioms, while their geometric properties arise from the parametrization

of group elements by points of a differentiable manifold. At the topological level, a Lie group is homogeneous, meaning that every point of the manifold parametrizing the group looks the same as any other point.

The dimension of a Lie group is the dimension of the manifold that parametrizes the group operations. If G is an r -dimensional smooth manifold, the corresponding Lie group is also called an r -parameter Lie group. In particular, every Lie group is a *topological group*, that is, a topological space equipped with a group structure such that the multiplication and inversion maps are continuous.

If G is a Lie group, any element $g \in G$ defines the maps $L_g, R_g: G \rightarrow G$, called respectively the *left translation* and *right translation*, given by

$$L_g(h) = gh, \quad R_g(h) = hg.$$

The maps L_g and R_g are smooth and, in fact, are diffeomorphisms of G .

Example 7. (Lie Groups). Each of the following manifolds is a Lie group with the specified group operation.

- (a) The general linear group $GL(n, \mathbb{R})$ is the set of all invertible $n \times n$ matrices with real entries. It is a group with group operation given by matrix multiplication, as already noted in Example 6(d), and it is an open subset of the vector space $\mathcal{M}(n, \mathbb{R})$. As will be shown below, multiplication is smooth since the entries of the matrix product AB are polynomials in the entries of A and B .
- (b) Let $GL^+(n, \mathbb{R})$ denote the subset of $GL(n, \mathbb{R})$ consisting of matrices with positive determinant. Since $\det(AB) = \det(A)\det(B)$ and $\det(A^{-1}) = 1/\det(A)$, it is a subgroup of $GL(n, \mathbb{R})$. Because this subset is the preimage of $(0, \infty)$ under the continuous determinant function, it is an open subset of $GL(n, \mathbb{R})$ and hence an n^2 -dimensional manifold. The group operations are the restrictions of those on $GL(n, \mathbb{R})$ and are therefore smooth. Thus $GL^+(n, \mathbb{R})$ is a Lie group.
- (c) The complex general linear group $GL(n, \mathbb{C})$ is the group of invertible complex $n \times n$ matrices under matrix multiplication. It is an open submanifold of $\mathcal{M}(n, \mathbb{C})$ and hence a $2n^2$ -dimensional smooth manifold. It is a Lie group since matrix multiplication and inversion are smooth functions of the real and imaginary parts of the matrix entries.
- (d) If V is an arbitrary real or complex vector space, $GL(V)$ denotes the set of invertible linear transformations from V to itself. It is a group under composition of functions. If V has finite dimension n , any choice of basis of V determines an isomorphism of $GL(V)$ with $GL(n, \mathbb{R})$ or $GL(n, \mathbb{C})$, so that $GL(V)$ is a Lie group. The transition map between two such isomorphisms is given by a map of the form $A \mapsto BAB^{-1}$, where B is the change-of-basis matrix, which is smooth. Consequently, the smooth manifold structure on $GL(V)$ is independent of the choice of basis.
- (e) The field of real numbers \mathbb{R} is a Lie group under addition, with the inverse given by $x \mapsto -x$.
- (f) Let $G = \mathbb{R}^r$ with its natural manifold structure, and let the group operation be vector addition $(x, y) \mapsto x + y$. The inverse of a vector x is $-x$. These operations are smooth, so \mathbb{R}^r provides an example of an r -parameter Abelian Lie group. Similarly, \mathbb{C} and \mathbb{C}^n are Lie groups under addition.
- (g) Let $G = SO(2)$ be the group of planar rotations. That is,

$$G = \left\{ \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} : 0 \leq \theta < 2\pi \right\},$$

where θ denotes the angle of rotation of a vector under multiplication by the rotation matrix. Note that G can be identified with the unit circle

$$\mathbb{S}^1 = \{(\cos \theta, \sin \theta) : 0 \leq \theta < 2\pi\}$$

in \mathbb{R}^2 , which allows one to endow $SO(2)$ with a manifold structure.

2.10.2. Matrix Groups

In this section we consider several matrix Lie groups that play an important role in matrix analysis and matrix computations. For brevity, in what follows we shall denote the general linear group $GL(n, \mathbb{R})$ simply by $GL(n)$.

- **General linear group**

As noted above, the general linear group

$$GL(n) = \{A \in \mathcal{M}(n) : \det(A) \neq 0\}$$

of all invertible $n \times n$ matrices with real entries is a smooth manifold of dimension n^2 , since it is an open subset of the space of all $n \times n$ matrices $\mathcal{M}(n)$. Indeed,

$$GL(n) = \mathbb{R}^{n^2} \setminus \{\det(A) = 0\},$$

where the space \mathbb{R}^{n^2} is identified with the space $\mathcal{M}(n)$ of all real $n \times n$ matrices. Since the determinant function $\det: \mathcal{M}(n) \rightarrow \mathbb{R}$ is continuous — because $\det(A)$ is a polynomial in the matrix entries — the set $\mathbb{R}^{n^2} \setminus \{\det(A) = 0\}$ is open. Hence it is a domain and therefore a smooth manifold of dimension n^2 .

To prove that $GL(n)$ is a Lie group, we must verify that matrix multiplication and matrix inversion are smooth operations. Given two matrices A and B in $GL(n)$, the element in position (i, j) of the product AB is given by

$$(AB)_{ij} = \sum_{k=1}^n a_{ik}b_{kj}.$$

Thus, $(AB)_{ij}$ is a polynomial of degree two in the entries of A and B . Consequently, the matrix multiplication map

$$m: GL(n) \times GL(n) \rightarrow GL(n)$$

is a smooth mapping.

Recall that the (ij) -minor of a matrix A is the determinant of the submatrix A_{ij} , obtained by deleting the i -th row and the j -th column of A . According to Cramer's rule the (ij) -th entry of A^{-1} is given by

$$(A^{-1})_{ij} = \frac{1}{\det(A)} \cdot (-1)^{i+j} (\text{the}(j, i) - \text{minor of } A),$$

which is a smooth function of the entries a_{ij} , provided that $\det(A) \neq 0$. That is, the mapping

$$\mathcal{M}(n) \rightarrow \mathbb{R}, A \mapsto (A^{-1})_{ij}$$

is smooth, since it depends smoothly on the entries of A . Therefore, the matrix inversion map

$$i: GL(n) \rightarrow GL(n)$$

is also smooth.

The complex general linear group $GL(n, \mathbb{C})$ is likewise a Lie group with respect to matrix multiplication and inversion. The set $GL(n, \mathbb{C})$ is an open subset of $\mathcal{M}(n, \mathbb{C})$ and hence is a smooth manifold of dimension $2n^2$. It is a Lie group because matrix multiplication and inversion are smooth functions of the real and imaginary parts of the matrix entries.

- **Special Linear Group**

This group is defined as

$$SL(n) = \{A \in \mathcal{M}(n) : \det(A) = 1\}.$$

Geometrically, $SL(n)$ consists of all transformations of \mathbb{R}^n that preserve both volume and orientation. It can be shown that $SL(n) = \det^{-1}(1)$ is a smooth manifold of dimension $n^2 - 1$, defined

by the condition $\det(A) - 1 = 0$. Since this manifold is a subset of the Lie group $GL(n)$ with the operation inherited from $GL(n)$, $SL(n)$ is also a Lie group. Moreover, the tangent space of $SL(n)$ at the identity is the subspace of $\mathcal{M}(n, \mathbb{R})$ consisting of all matrices with zero trace.

- **Group of Orthogonal Matrices**

The group of orthogonal $n \times n$ matrices is defined as

$$O(n) = \{A \in GL(n) : A^T A = I_n\}.$$

Thus, $O(n)$ is a subset of \mathbb{R}^{n^2} , defined by n^2 equations

$$A^T A - I_n = 0$$

in terms of the entries a_{ij} of the matrix A . It can be shown that exactly $n(n+1)/2$ of these equations, corresponding to the entries on and above the diagonal, are independent and satisfy the maximal rank condition for $O(n)$ everywhere. Therefore, $O(n)$ is a submanifold of $GL(n)$ of dimension $n - n(n+1)/2 = n(n-1)/2$. Moreover, matrix multiplication and the matrix inversion operation remain smooth when restricted to $O(n)$. Consequently, $O(n)$ itself is a Lie group.

- **Special Orthogonal Group**

The equation

$$A^T A = I_n,$$

used in the definition of the orthogonal group $O(n)$, in particular implies that every matrix $A \in O(n)$ is invertible with $A^{-1} = A^T$. Consequently, the determinant of $A \in O(n)$ must satisfy $(\det(A))^2 = 1$, i.e., $\det(A) = \pm 1$. In this way, $O(n)$ is divided into two disconnected components: the subset of matrices with determinant $+1$ and the subset of matrices with determinant -1 .

If A and B both have determinant -1 , then their product AB has determinant $+1$. Therefore, the subset of orthogonal matrices with determinant -1 is not closed under multiplication and is not a subgroup of $O(n)$. The other component, however, is a Lie subgroup of $O(n)$ and is called the *special orthogonal group*, denoted by $SO(n)$,

$$SO(n) = \{A \in O(n) : \det(A) = 1\} \subset O(n).$$

The subgroup $SO(n)$ is a Lie group.

- **Unitary and Special Unitary Groups**

The *unitary group* $U(n)$ is defined as

$$U(n) := \{A \in GL(n, \mathbb{C}) : A^H A = I_n\},$$

where $A^H = \bar{A}^T$ denotes the Hermitian (complex conjugate transpose) of A . A similar argument as for $O(n)$ shows that $U(n)$ is a submanifold of $GL(n, \mathbb{C})$ and that $\dim U(n) = n^2$.

The *special unitary group* $SU(n)$ is defined as the subgroup of $U(n)$ consisting of matrices with determinant equal to 1.

- **Group of Upper Unit Triangular Matrices**

The group $T(n)$ of upper triangular matrices with ones on the main diagonal is an $n(n-1)/2$ -parameter Lie group. As a manifold, $T(n)$ can be identified with the Euclidean space $\mathbb{R}^{n(n-1)/2}$, since each matrix is uniquely determined by its entries above the diagonal. For example, in the case of $T(3)$ we identify the matrix

$$\begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix} \in T(3)$$

with the vector (x, y, z) in \mathbb{R}^3 . However, except for the special case $T(2)$, the group $T(n)$ is not isomorphic to the Abelian Lie group $\mathbb{R}^{n(n-1)/2}$.

Several other important matrix groups are defined by imposing linear or quadratic constraints on the n^2 entries of $GL(n, \mathbb{F})$ or $SL(n, \mathbb{F})$.

3. Geometry of Matrix Space

3.1. The Matrix Space

Let us consider the linear operator

$$\mathcal{A} : \mathbb{F}^n \rightarrow \mathbb{F}^m,$$

represented by a rectangular matrix $A \in \mathbb{F}^{m \times n}$. The finite-dimensional space of all linear operators from \mathbb{F}^n to \mathbb{F}^m is isometric to the vector space $\mathbb{F}^{m \cdot n}$, that is, the two spaces may be regarded as the same metric space. This makes it possible to study the space of linear operators using methods from the analysis of metric and topological spaces.

Let the set $\mathcal{M}(m, n)$ of all $m \times n$ matrices with entries in \mathbb{F} be endowed with the topology induced by the *natural mapping* $i : \mathbb{F}^{m \cdot n} \rightarrow \mathcal{M}(m, n)$:

$$(x_1, x_2, \dots, x_{mn}) \mapsto \begin{bmatrix} x_1 & x_2 & \dots & x_n \\ x_{n+1} & x_{n+2} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{(m-1)n+1} & x_{(m-1)n+2} & \dots & x_{mn} \end{bmatrix}.$$

The mapping i is a homeomorphism that induces on $\mathcal{M}(m, n)$ the structure of a smooth (C^∞) manifold of dimension mn . In other words, the set $\mathcal{M}(m, n)$ of all rectangular matrices $A \in \mathbb{F}^{m \times n}$ forms a smooth manifold of dimension $m \cdot n$ embedded in the vector space $\mathbb{F}^{m \cdot n}$.

According to the above considerations, a complex rectangular matrix $A \in \mathbb{C}^{m \times n}$ can be represented as a point in the complex linear space $\mathbb{C}^{m \cdot n}$ of dimension $m \cdot n$. Similarly, a real $m \times n$ matrix A can be viewed as a point in the real space \mathbb{R}^N , with $N = m \cdot n$, by arranging its entries in a prescribed order and interpreting them as coordinates. The corresponding space is called the *matrix space* and is isometric to a vector space of the same dimension. In the same way, a complex polynomial of degree n can be identified with a point in \mathbb{C}^N , where $N = n + 1$ by using its coefficients as coordinates.

The *distance between two points in the matrix space* is defined by

$$\text{dist}(A, B) = \|A - B\|.$$

Depending on the norm used, we shall implement the 2-norm distance $\text{dist}_2(A, B) = \|A - B\|_2$ or the Frobenius norm distance $\text{dist}_F(A, B) = \|A - B\|_F$.

3.2. Generic and Well-Posed Problems

Consider the following idealized geometric interpretation of a computational problem $y = f(x)$. Let \mathcal{X} denote the space of the data x , let \mathcal{Y} denote the space of solutions y and let

$$f : \mathcal{X} \rightarrow \mathcal{Y}$$

be (a generally nonlinear) operator mapping $x \in \mathcal{X}$ to $y \in \mathcal{Y}$. In general, the spaces \mathcal{X} and \mathcal{Y} may be arbitrary topological spaces, but for the purposes of our analysis we restrict attention to metric spaces and, in particular, to normed vector spaces.

The components of x are called *parameters* and the space \mathcal{X} is referred to as the *parameter space*. The *dimension* of the parameter space denoted by $\dim(\mathcal{X})$ is equal to the number N of parameters. Each computational problem is therefore identified with a point in \mathcal{X} .

If the vector x represents a $m \times n$ matrix with independent entries, then \mathcal{X} is the full matrix space with dimension $m \cdot n$. However, if the number of parameters N is smaller than $m \cdot n$, the matrix entries are no longer independent. Consequently, the parameter space in this case will have smaller dimension than the corresponding full matrix space.

If the mapping $f : x \rightarrow y$ is linear, the problem is called *linear*; otherwise it is called *nonlinear*. If the inverse operator f^{-1} exists (at least locally) for the given data x , the problem is said to be *regular*. In the opposite case, the problem is called *singular*.

Suppose $\mathcal{P}(x)$ is a property that may be asserted about a problem x . This property represents a function $\mathcal{P} : \mathcal{X} \rightarrow \{0, 1\}$, where $\mathcal{P}(x) = 1$ (resp. 0) indicates that \mathcal{P} holds (resp. fails) at x . In applications where x represents a data of a physical problem subject to errors and uncertainties, it is important to understand the topological features of \mathcal{P} .

For instance, if \mathcal{P} holds at a nominal parameter value x_0 , it is useful to know whether \mathcal{P} also holds at points $x = x_0 + \delta x$ in a neighborhood of x_0 , corresponding to small deviations δx of the parameters from their nominal values. Usually, the property \mathcal{P} of interest holds for all sets of parameter values except those corresponding to points x lying on some surface in the parameter space, which are thus atypical.

From the point of view of algebraic geometry [34], such a surface represents a *variety* \mathcal{S} whose dimension satisfies $0 < \dim(\mathcal{S}) < N$. We call \mathcal{S} the *variety of singular cases*. Typically, \mathcal{S} is a closed subset of \mathcal{X} , while the set of problems for which \mathcal{P} holds contains an open and dense subset of \mathcal{X} .

We say that a property \mathcal{P} is *generic relative to* \mathcal{S} if $\mathcal{P}(x) = 0$ only for points $x \in \mathcal{S}$. A generic property of a parameter space is a property that holds at “almost all” points of that space. Intuitively, a generic property is one that “almost all” problems satisfy. Equivalently, if the parameters are chosen at random, then $\mathcal{P} = 1$ with probability 1.

It should be emphasized that, depending on the problem, the variety of singular cases may have a complex structure. In the special case when the variety has no singular points, it is referred to as manifold.

Since regular problems are generic, they have been studied more thoroughly than the corresponding singular problems. Therefore, the analysis of generic cases is always a primary focus when investigating phenomena and processes described by a given mathematical model. Nevertheless, there are situations in which it becomes necessary to examine non-generic cases. In this context, it is appropriate to quote the words of the great German mathematician Leopold Kronecker, spoken in 1874 [35]: *“It is customary—especially in algebraic problems—to encounter truly new difficulties when one moves away from the cases usually regarded as general. As soon as one penetrates beneath the surface of the so-called generality, which excludes all particularities in the true generality encompassing all singularities, one typically first meets the real challenges of investigation, but at the same time also the richness of new perspectives and phenomena contained in its depths.”*

A property \mathcal{P} is said to be *well posed at* x if \mathcal{P} holds through some neighborhood of x in \mathcal{X} . A problem parametrized by data in \mathcal{X} is said to be well posed at a point x_0 , if it is solvable for all data points $x_0 + \delta$ in some neighborhood of x_0 . According to the French mathematician Jacques Hadamard [36], a mathematical problem is well posed, if it satisfies the following three conditions:

1. a solution exists,
2. the solution is unique,
3. the solution depends continuously on the data of the problem.

If a problem fails to satisfy at least one of these conditions – namely, if a solution does not exist for some set \mathcal{S} of parameters $x \in \mathcal{S}$, if the solution is not unique, or if the solution does not depend continuously on the parameters – then the problem is called an *ill-posed problem*.

The ill-posed problems typically form a variety \mathcal{S} within the space of all problems. This variety is referred to as the *set of ill-posed problems*.

If the property \mathcal{P} is generic relative to \mathcal{S} , then \mathcal{P} is well posed at every point in the complement \mathcal{S}^c (where $\mathcal{S} \cup \mathcal{S}^c = \mathcal{X}$). Since the dimension of \mathcal{S} is less than the dimension of \mathcal{X} , it follows that almost all problems in \mathcal{X} are well posed.

It should be noted, however, that many meaningful and important problems in physics and computational mathematics are not well-posed, that is, they are ill-posed.

Example 8. Consider the problem of finding the inverse of a square matrix A of order n . In this case, the property \mathcal{P} is the nonsingularity of the matrix A and the set \mathcal{S} consists of all singular matrices. The set of singular matrices is an algebraic variety, since it can be defined as the zero set of the polynomial $\det(A)$. Thus, the equation $\det(A) = 0$ describes the variety \mathcal{S} .

For example, if $n = 2$, $N = 3$ and

$$A = \begin{bmatrix} a & 1 \\ c & b \end{bmatrix},$$

then the variety \mathcal{S} is determined by the equality $ab = c$ and represents a hyperbolic paraboloid – a two dimensional manifold in the three-dimensional parameter space (see Figure 23). All points outside \mathcal{S} correspond to nonsingular matrices A , that is matrices for which $\mathcal{P}(A) = 1$.

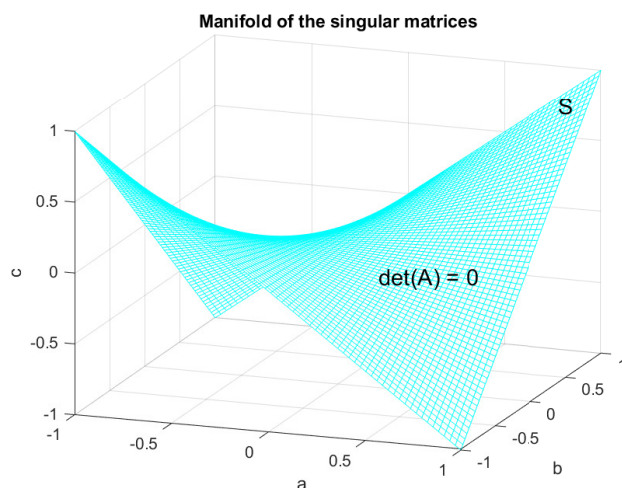


Figure 23. Variety of 2×2 singular matrices in the parameter space.

Example 9. Consider the eigenvalue problem for the same matrix, as in the previous example. In this case, the property \mathcal{P} is possessed by matrices with distinct eigenvalues, and the set \mathcal{S} consists of all matrices with a double eigenvalue. Since the eigenvalues of A are given by

$$\lambda_{1,2} = \frac{a + b \pm \sqrt{(a + b)^2 - 4(ab + c)}}{2},$$

the set \mathcal{S} is obtained by setting the discriminant equal to zero:

$$(a + b)^2 - 4(ab + c) = 0.$$

The corresponding surface in the three-dimensional parameter space is shown in Figure 24.

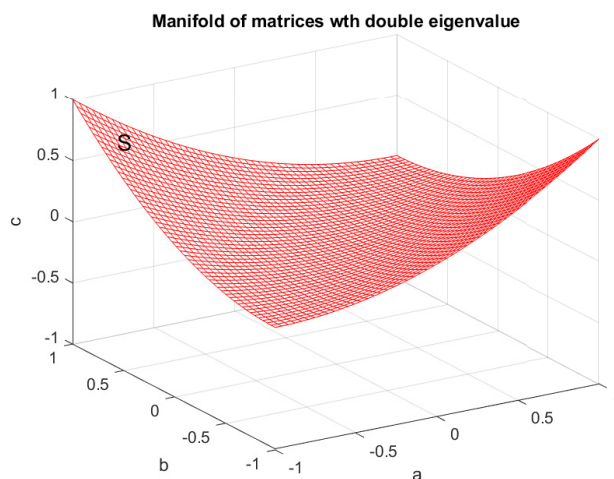


Figure 24. Variety of 2×2 matrices with a double eigenvalue.

3.3. Conditioning of Computational Problems

To obtain an accurate solution of a computational problem, it is not sufficient for the problem to be well posed. If the problem lies close, in some sense, to the set S of ill-posed problems, then it may be highly sensitive to variations in the data, and the discrepancy between the computed and the exact solutions can be very large.

The *sensitivity* of a given mathematical problem describes how its solution varies in the vicinity of a nominal solution. This important property of mathematical problems is the subject of *perturbation theory*, which is widely used in science and engineering.

A simplified quantitative characterization of sensitivity can be obtained by using the notion of *conditioning*.

Conditioning is a property of a computational problem that characterizes the sensitivity of its solution to small changes (or *perturbations*) in the data. A problem is said to be *well conditioned*, if small perturbations in the data lead to small changes in the solution. Conversely, if small changes in the data may result large changes in the solution, the problem is called *ill conditioned*.

It should be emphasized that the conditioning is an intrinsic property of the problem itself. It does not depend on the numerical precision used to solve the problem, nor does it depend on the particular numerical method implemented.

The theory of conditioning is developed in [37] and [38]. Various kinds of condition numbers are discussed in [39] and [40], while the computing of condition estimates is addressed in [42], Ch. 15, and [41,43,44].

3.4. Condition Numbers

Consider the computational problem $y = f(x)$, where f maps the input data x to the output y . The problem is said to be *well conditioned* if $f(x)$ is relatively insensitive to small perturbations δx in x , that is, $f(x)$ remains close to $f(x_0)$ when $x = x_0 + \delta x$ is near x_0 . The precise meaning of “insensitive”, “small” and “close” depends on the specific context of the problem. Conversely, if small changes in x produce large variations in f , the problem is ill conditioned.

The conditioning of a computational problem can be formally characterized as illustrated in Figure 25 where $\|\cdot\|$ denotes the vector norms in \mathcal{X} and \mathcal{Y} , respectively.

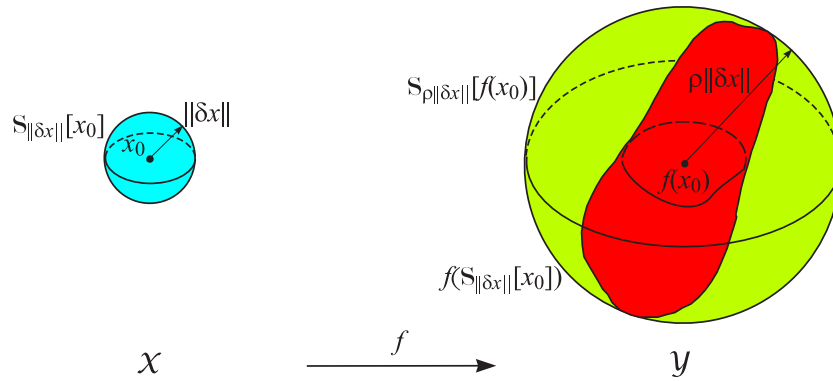


Figure 25. Geometric interpretation of the condition number.

Definition 14. (Absolute condition number). The *absolute δx -condition number* of the mapping $f : \mathcal{X} \rightarrow \mathcal{Y}$ at the point $x_0 \in \mathcal{X}$ is defined as

$$\kappa_{\delta x}^{abs}(f, x_0) = \inf\{\rho : f[\mathbb{S}_x(x_0, \|\delta x\|)] \subset \mathbb{S}_y(f(x_0), \rho\|\delta x\|)\}. \quad (3)$$

According to this definition, the condition number $\kappa_{\delta x}^{abs}$ is the smallest number ρ such that the image $f[\mathbb{S}_x(x_0, \|\delta x\|)]$ of the sphere $\mathbb{S}_x(x_0, \|\delta x\|)$ lies entirely within the sphere $\mathbb{S}_y(f(x_0), \rho\|\delta x\|)$ of a radius $\rho\|\delta x\|$.

Defined in this way, the condition number can be interpreted as a *Lipschitz constant* for the mapping f , i.e., it is the smallest number M for which

$$\|f(x_0 + \delta x) - f(x_0)\| \leq M\|x - x_0\|$$

holds for all points on the boundary of $\mathbb{S}(x_0, \|\delta x\|)$.

As illustrated in Figure 25, the condition number κ provides a quantitative measure of how perturbations in the data are transmitted to the solution space \mathcal{Y} . If κ is relatively small, the problem is *well conditioned*; if κ is large, the problem is *ill conditioned*.

Definition 15. (Relative condition number). The *relative δx -condition number* of the mapping $f : \mathcal{X} \rightarrow \mathcal{Y}$ at the point $x_0 \in \mathcal{X}$ is defined as

$$\kappa_{\delta x}^{rel}(f, x_0) = \inf\{\rho : f[\mathbb{S}_x^r(x_0, \|\delta x\|)] \subset \mathbb{S}_y^r(f(x_0), \rho\|\delta x\|)\}. \quad (4)$$

In many practical applications, it is common to consider the condition number in the limit of infinitesimally small perturbations $\|\delta x\|$. Although this represents a theoretical approximation, it is widely used in perturbation analysis because it is easier to compute.

Definition 16. (Asymptotic condition number). The *asymptotic absolute and asymptotic relative condition number* of $f : \mathcal{X} \rightarrow \mathcal{Y}$ are defined, respectively, as

$$(i) \kappa^{abs}(f, x_0) = \lim_{\|\delta x\| \rightarrow 0} \kappa_{\delta x}^{abs}(f, x_0), \quad (5)$$

$$(ii) \kappa^{rel}(f, x_0) = \lim_{\|\delta x\| \rightarrow 0} \kappa_{\delta x}^{rel}(f, x_0). \quad (6)$$

If the Jacobian matrix $D_{x_0}f$ of the mapping f exists, then the asymptotic absolute and relative condition numbers are given by

$$(i) \kappa^{abs}(f, x_0) = \|D_{x_0}f\|, \quad (7)$$

$$(ii) \kappa^{rel}(f, x_0) = \frac{\|D_{x_0}f\| \|x_0\|}{\|f(x_0)\|}, \quad (8)$$

where $\|D_{x_0}f\|$ denotes the operator (subordinate norm) of the derivative $D_{x_0}f$ in \mathcal{Y} .

The asymptotic condition number is usually referred to as the *condition number of the problem*. Note that this condition number does not exist for all mappings f .

The asymptotic absolute and relative condition numbers can be expressed as

$$\kappa^{abs}(f, x_0) := \limsup_{\|\delta x\| \rightarrow 0} \frac{\|f(x_0 + \delta x) - f(x_0)\|}{\|\delta x\|} = \|D_{x_0}f\|, \quad (9)$$

$$\begin{aligned} \kappa^{rel}(f, x_0) &:= \limsup_{\|\delta x\| \rightarrow 0} \frac{\|f(x_0 + \delta x) - f(x_0)\| / \|f(x_0)\|}{\|\delta x\| / \|x_0\|} \\ &= \frac{\|D_{x_0}f\| \|x_0\|}{\|f(x_0)\|} \end{aligned} \quad (10)$$

which coincides with the expressions in (7), (8).

From this definition, it follows that

$$\frac{\|f(x_0 + \delta x) - f(x_0)\|}{\|f(x_0)\|} \leq \kappa^{rel}(f, x_0) \frac{\|\delta x\|}{\|x_0\|} + o(\delta x), \quad (11)$$

where $h = o(|\delta x|)$ means $|h|/|\delta x| \rightarrow 0$ as $|\delta x| \rightarrow 0$.

Equation (11) shows that the relative change in the result can be approximated by the product of the relative condition number $\kappa^{rel}(f, x_0)$ and the relative perturbation $\|\delta x\|/\|x_0\|$ in the data. The ratio $\|\delta x\|/\|x_0\|$ is called *backward error* since the change in the solution is represented as an equivalent perturbation in the input data equivalent to the observed change in the solution. Computational methods for which the backward error is small are called *numerically stable*.

Consequently, the quantity

$$\kappa^{rel}(f, x_0) \frac{\|\delta x\|}{\|x_0\|}$$

provides an estimate of the relative *forward error* in the solution, given the relative backward error. If the condition number is large, even small perturbations in the input may produce a large forward error, highlighting the difficulty of the computational problem.

It should be noted that even well-conditioned problems may have specific perturbations for which the sensitivity estimate overstates the true change in the solution. When the sensitivity estimate consistently produces overly pessimistic bounds, it is called a *conservative estimate*.

3.5. Conditioning of Basic Matrix Problems

In this section we present two case studies that illustrate the concept of conditioning in matrix computations.

3.5.1. Conditioning of a Linear System of Equations

Consider the linear system

$$Ay = x,$$

where $A \in \mathbb{C}^{n \times n}$ is nonsingular and $y = A^{-1}x$, $x \neq 0$. Solving the system corresponds to applying the linear operator $x \rightarrow A^{-1}x$.

The entries of A serve as parameters, and the set of singular matrices forms a variety \mathcal{S} , representing ill-posed problems. Small perturbations δA may therefore lead to large changes in the solution.

Assuming that A is perturbed to $A + \delta A$, the perturbed solution satisfies

$$y + \delta y = (A + \delta A)^{-1}x, \quad (12)$$

The matrix $A + \delta A$ remains nonsingular provided

$$\|A^{-1}\delta A\| < 1. \quad (13)$$

Under this condition, standard perturbation theory [45], Ch. 3, [46], Sect. 2.6, [47], Sect. 1.2) yields the relative error bound

$$\frac{\|\delta y\|}{\|y\|} \leq \frac{\|A\|\|A^{-1}\|}{1 - \|A^{-1}\delta A\|} \frac{\|\delta A\|}{\|A\|}. \quad (14)$$

According to Definition 15, the product

$$\kappa^{rel} = \|A\|\|A^{-1}\|$$

represents the *relative condition number* of the linear system. It satisfies $\kappa^{rel} \geq 1$ and is invariant under scalar multiplication of A . For sufficiently small perturbations, the bound simplifies asymptotically to

$$\frac{\|\delta y\|}{\|y\|} \leq \kappa^{rel} \frac{\|\delta A\|}{\|A\|}. \quad (15)$$

If κ^{rel} is small, the system is *well conditioned*; if large, it is *ill conditioned*. When $\kappa^{rel} \approx 1/\mathbf{u}$, where \mathbf{u} is the unit roundoff, the matrix is considered *singular to working precision*, and accurate solutions cannot be expected numerically.

Following standard notation, the condition number is denoted by

$$\text{cond}(A) = \|A\|\|A^{-1}\|.$$

The value of the condition number depends on the chosen norm. In particular, using the singular value decomposition $A = U\Sigma V^T$,

$$\text{cond}_2(A) = \frac{\sigma_{\max}}{\sigma_{\min}},$$

where σ_{\max} and σ_{\min} are the largest and smallest singular values of A . The 2–norm or the Frobenius–norm condition numbers are invariant under unitary (orthogonal) transformations.

Example 10. Consider the linear system $Ay = x$, where

$$A = \begin{bmatrix} -6 & 2 & 5 & -3 & p & 1 \\ 3 & -1 & -4 & 3 & -8 & -2 \\ q & 7 & -2 & -5 & 4 & -3 \\ -2 & -8 & 9 & 4 & -5 & 7 \\ 9 & 0 & 6 & -1 & -7 & 3 \\ 5 & 8 & -3 & -6 & 2 & 4 \end{bmatrix}, \quad x = \begin{bmatrix} -2 \\ 7 \\ -5 \\ -3 \\ 9 \\ 1 \end{bmatrix}$$

and $p \in [-10, 10]$, $q \in [-10, 10]$ are variable parameters. Assume that the matrix A is subject to the perturbation

$$\delta A = 10^{-6} \begin{bmatrix} 2 & 5 & 8 & 9 & 3 & 4 \\ 6 & -4 & -7 & -2 & -5 & 1 \\ -3 & 1 & 0 & 4 & 6 & -2 \\ -4 & 1 & 2 & -9 & -8 & 5 \\ 8 & -5 & 3 & 2 & 0 & 1 \\ -7 & 9 & -2 & -3 & 5 & 0 \end{bmatrix}.$$

In Figure 26, we display the relative perturbation

$$\frac{\|\delta y\|_2}{\|y\|_2}$$

in the solution caused by the perturbation δA , as a function of the parameters p and q . This is shown together with the corresponding relative perturbation estimate

$$\frac{\|\delta y^{est}\|_2}{\|y\|_2} = \frac{\text{cond}_2(A) \frac{\|\delta A\|_2}{\|A\|_2}}{1 - \text{cond}_2(A) \frac{\|\delta A\|_2}{\|A\|_2}}.$$

The peak values in the Figure represent ill conditioned matrices. The maximum condition number of the perturbed matrices is $\text{cond}_2(A) = 3.21 \times 10^5$ which leads to maximum relative perturbations of size 0.1.

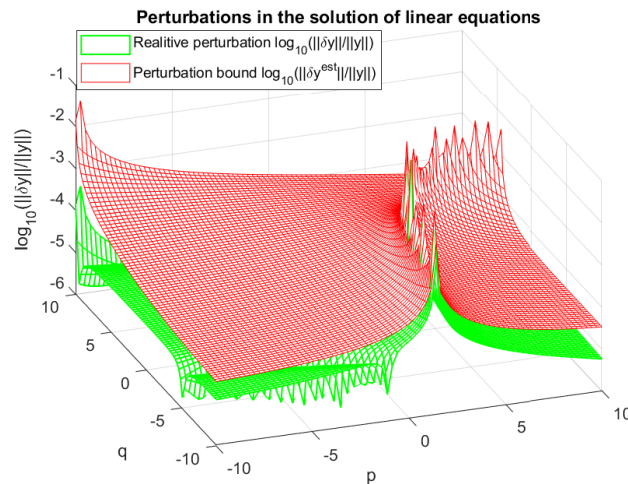


Figure 26. Perturbations in the solution of a linear system in the parameter space.

3.5.2. Conditioning of the Eigenvalue Problem

We consider, in a simplified setting, the problem of sensitivity of the eigenvalues of a matrix. This problem is characterized by fundamentally different properties in the regular case, when the eigenvalues are distinct, and in the singular case, when the eigenvalues are multiple and possess nonlinear elementary divisors.

Let us first examine the asymptotic sensitivity of the eigenvalue problem in the regular case. In this case, the eigenvalue problem is well posed, since the variations in the eigenvalues depend linearly on the perturbations in the entries of the matrix.

Let $A \in \mathbb{C}^{n \times n}$ be a given square matrix. If A has a simple eigenvalue λ , with corresponding right eigenvector $x \in \mathbb{C}^n$ and left eigenvector $y \in \mathbb{C}^n$, then

$$Ax = \lambda x \quad \text{and} \quad y^H A = \lambda y^H,$$

where y^H denotes the conjugate transpose of y . Since λ a simple eigenvalue, it follows that

$$y^H x = \langle x, y \rangle \neq 0.$$

For every sufficiently small perturbation $\delta A \in \mathbb{C}^{n \times n}$, there exists a unique eigenvalue $\lambda + \delta \lambda$ of $A + \delta A$, that is close to λ . Therefore, we have

$$(A + \delta A)(x + \delta x) = (\lambda + \delta \lambda)(x + \delta x),$$

which implies that, up to terms of second order,

$$\delta Ax + A\delta x \approx \delta \lambda x + \lambda \delta x.$$

Multiplying from the left by y^H , we obtain

$$\delta\lambda = \frac{1}{\langle x, y \rangle} y^H \delta A x + o(\|\delta A\|).$$

Moreover,

$$\max_{\|\delta A\|_F \leq 1} |y^H \delta A x| = \|x\| \|y\|.$$

Therefore, *absolute condition number* of the matrix A with respect to the regular eigenvalue problem can be defined as

$$\kappa_\lambda^{abs}(A) := \frac{\|x\| \|y\|}{\langle x, y \rangle}.$$

We have that

$$\kappa_\lambda^{abs}(A) = \|P\|_F, \quad (16)$$

where

$$P = \frac{xy^H}{y^H x}$$

is the *spectral projector* onto the invariant subspace generated by x . If the right and left eigenvector are normalized so that $y^H x = 1$, then

$$\kappa_\lambda^{abs}(A) := \|x\| \|y\|.$$

The condition number is homogeneous, since multiplying A by a scalar does not change $\kappa_\lambda^{abs}(A)$. It is also invariant under unitary (or orthogonal) transformations.

If $\kappa_\lambda^{abs}(A)$ is large, then λ is poorly conditioned. Poorly conditioned eigenvalues are computed with large errors as a consequence of their high sensitivity to perturbations in the matrix. As the separation between eigenvalues decreases, their sensitivity increases, and in the limiting case of defective eigenvalues, their condition numbers become infinitely large. In such cases, the linear estimate

$$|\delta\lambda| \leq \kappa_\lambda^{abs}(A) \|\delta A\|, \quad (17)$$

is no longer valid, and the eigenvalue problem becomes ill-posed. In this case, it is justified to set $\kappa_\lambda^{abs}(A) := \infty$, where λ is a multiple eigenvalue of A .

We note that the eigenvalue sensitivity estimates provided by the linear algebra package LAPACK [48] and the software system MATLAB[®] [3] yield meaningful results only for well-posed problems, specifically when the matrices have simple eigenvalues, i.e., eigenvalues to which correspond linearly independent eigenvectors.

Example 11. Consider the upper triangular matrix

$$A = \begin{bmatrix} \lambda_1 & 1 \\ 0 & \lambda_1 + \varepsilon \end{bmatrix}, \quad \varepsilon \neq 0.$$

The matrix A has two simple eigenvalues λ_1 and $\lambda_2 = \lambda_1 + \varepsilon$. The corresponding spectral projectors are

$$P_{\lambda_1} = \begin{bmatrix} 1 & -1/\varepsilon \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad P_{\lambda_2} = \begin{bmatrix} 0 & 1/\varepsilon \\ 0 & 1 \end{bmatrix}.$$

Both eigenvalues have the same absolute condition number given by

$$\kappa_\lambda^{abs}(A) = \frac{\sqrt{1 + |\varepsilon|^2}}{|\varepsilon|}.$$

As $\varepsilon \rightarrow 0$, the condition number tends to infinity, since the two eigenvalues of A coalesce ($\lambda_1 = \lambda_2 = \lambda$) and the eigenvalue problem becomes ill-posed.

Since the eigenvalue condition number becomes infinite when λ is a multiple eigenvalue, the variety of singular cases for the eigenvalue problem consists of all matrices with multiple eigenvalues. This variety forms a hypersurface in the space of $n \times n$ matrices, i.e., it has dimension $N - 1$, where $N = n^2$.

For matrices with multiple eigenvalues, small perturbations of the entries can cause large changes in the eigenvalues. Such matrices are characterized by eigenvalues that appear in nonlinear elementary divisors.

Following [49,50], we now examine the sensitivity of the eigenvalue problem in the singular case.

Let

$$J = \text{diag}(\Gamma_1^1, \dots, \Gamma_1^{r_1}, \dots, \Gamma_q^1, \dots, \Gamma_q^{r_q}),$$

where, for $j = 1, \dots, q$,

$$\Gamma_j^1 = \dots = \Gamma_j^{r_j} = \begin{bmatrix} \lambda & 1 & & & \\ & \cdot & \cdot & & \\ & & \cdot & \cdot & \\ & & & \cdot & 1 \\ & & & & \lambda \end{bmatrix}$$

is a Jordan block of dimension n_j , repeated r_j times, and

$$n_1 \geq n_2 \geq \dots \geq n_q.$$

The eigenvalue λ is *semisimple* (participating in a scalar Jordan block, i.e., nondefective) if $q = n_1 = 1$, and *nonderogatory* (participating in only one Jordan block) if $q = r_1 = 1$. It follows that the algebraic and geometric multiplicities of λ , are, respectively,

$$m = \sum_{j=1}^q r_j n_j \quad \text{and} \quad g = \sum_{j=1}^q r_j.$$

Denote by x_1^k the right eigenvector and by y_1^k the left eigenvector of A associated with $\Gamma_1^k, k = 1, \dots, r_1$. With these eigenvectors we construct the matrices

$$X_1 = [x_1^1, \dots, x_1^{r_1}], \quad Y_1 = [y_1^1, \dots, y_1^{r_1}]$$

corresponding to the Jordan blocks of maximum size n_1 . Note that the r_1 columns of X_1 and Y_1 are linearly independent right and left eigenvectors, respectively, each associated with a separate Jordan chain of maximal length n_1 .

Assume that the matrix A is perturbed to $A + \varepsilon B$, where ε is small. It is shown in [49,50] that the eigenvalues $\tilde{\lambda}$ of $A + \varepsilon B$ converging to λ as $\varepsilon \rightarrow 0$ satisfy

$$|\tilde{\lambda} - \lambda| \leq c(\alpha\varepsilon)^{1/n_1}, \quad (18)$$

for all sufficiently small positive ε , where $\alpha = \|X_1 Y_1^H\|_2$ and $c > 1$.

The bound (18) shows that the sensitivity of a multiple eigenvalue depends on ε^{1/n_1} , where n_1 is the size of the largest Jordan block associated with λ . In the case where λ is simple, we have $n_1 = r_1 = 1$, and $\alpha = \|x\| \|y\|$, where the right eigenvector x and the left eigenvector y are normalized so that $y^H x = 1$. In this case, the bound (18) coincides with the bound (17) valid for a simple eigenvalue. It is important to note that the bound (18) remains finite, while the linear bound (17) becomes infinite in this case.

If λ is nonderogatory (i.e., there is only one Jordan block corresponding to λ), then $r_1 = 1$, and again in (18) we obtain $\alpha = \|x\|\|y\|$, where x and y are the corresponding right and left eigenvectors associated with Γ_1^1 .

The following example illustrates how the dimension of the Jordan block associated with a multiple eigenvalue affects its sensitivity.

Example 12. Consider a Jordan block $J_n(\lambda)$ of n th order with eigenvalue λ . If the zero entry in position $(n, 1)$ is replaced by a small number $\varepsilon > 0$, the characteristic equation of the perturbed block becomes

$$(\lambda - \tilde{\lambda})^n - (-1)^n \varepsilon = 0$$

and the multiple eigenvalue λ is split into n distinct eigenvalues

$$\tilde{\lambda}_k = \lambda + \varepsilon^{1/n}(\cos(2k\pi/n) + j_o \sin(2k\pi/n)), \quad k = 0, 1, \dots, n-1.$$

In the given case $n_1 = 6, r_1 = 1$ and $J_n(\lambda)$ has one right eigenvector $x = [1 \ 0 \ 0 \ 0 \ 0 \ 0]^T$ and one left eigenvector $y = [0 \ 0 \ 0 \ 0 \ 1]^T$ so that $\alpha = \|x\|\|y\| = 1$. The perturbed eigenvalues satisfy

$$|\tilde{\lambda}_k - \lambda| = \varepsilon^{1/n},$$

as predicted by (18).

For instance, consider

$$J_6(-1) = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix},$$

with $n = 6$ and $\lambda = -1$. For $\varepsilon = 10^{-6}$, the perturbed eigenvalues are

$$\begin{aligned} & -0.9000000000000000, \\ & -0.9500000000000000 + j_o 0.086602540378444, \\ & -0.9500000000000000 - j_o 0.086602540378444, \\ & -1.0500000000000000 + j_o 0.086602540378444, \\ & -1.0500000000000000 - j_o 0.086602540378444, \\ & -1.1000000000000000. \end{aligned}$$

The perturbations in the eigenvalues are exactly $\varepsilon^{1/6} = 0.1$, which is 10^5 larger than the original perturbation $\varepsilon = 10^{-6}$.

In Figure 27, we show the eigenvalue perturbations of $J_6(-1)$ for 20 equally spaced values of ε between 10^{-8} (bottom circle) and 10^{-6} (top circle). Note that the number of the eigenvalue loci is equal to the size of the Jordan block.

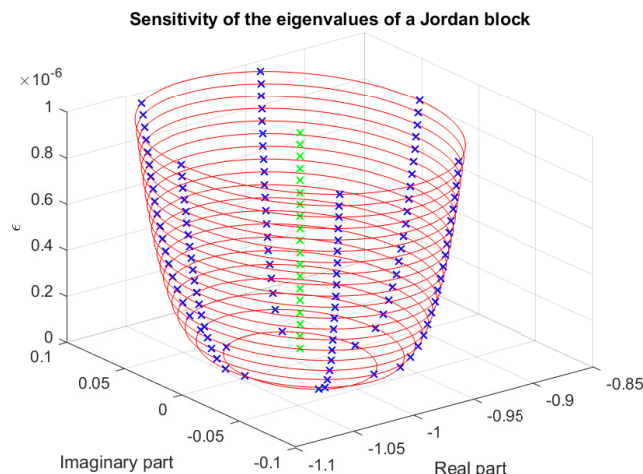


Figure 27. Sensitivity of the eigenvalues of a Jordan block for values of ϵ between 10^{-8} and 10^{-6} .

Thus, a perturbation of size ϵ can induce changes of order $\epsilon^{1/n}$ in the eigenvalues, which can be large even for moderate n . Since the trace of the matrix is unchanged, the mean of the eigenvalues $\sum_1^n \lambda_i / n$ remains λ .

In this example, the eigenvalues are highly sensitive to perturbations because the matrix is completely defective. In contrast, eigenvalues corresponding to linear elementary divisors generally exhibit low sensitivity. This highlights that the analysis of eigenvalue sensitivity must explicitly or implicitly account for the Jordan structure of the matrix.

A large body of results on the sensitivity of matrix eigenvalues can be found in the books [46,51–53], and [54]. The book by Stewart and Sun [55] provides a detailed presentation of various methods for perturbation analysis of eigenvalue and eigenvector problems. Comprehensive surveys of such methods are also given in [56] and [57]. Several important results on eigenvalue sensitivity have been published in [58–65], among others. Theoretical and practical aspects of eigenvalue conditioning analysis are addressed in [66–71].

3.6. Distance to an Ill-Posed Problem and Conditioning

The study of the sensitivity of a system of linear equations done in sect. 3.5, shows that with the increasing of the condition number the matrix becomes closer and closer to a singular matrix. The analysis of several other problems of numerical analysis shows that, in a similar way, the corresponding condition number is inversely proportional to the distance of the problem to the set of ill-posed problems. Thus, as a problem gets closer to the set of ill-posed ones, its condition number approaches infinity.

The geometry of ill-conditioning in numerical computations has been developed by Smale [72–74], Renegar [75,76], and Demmel [77–79]. These works show that many problems in numerical analysis, particularly in matrix computations, satisfy the property that the condition number of a problem is proportional to (or bounded by a multiple of) the reciprocal of the distance to the set of ill-posed problems. Consequently, as a problem approaches the variety of ill-posed problems, its condition number can increase without bound.

Determining the probability of encountering problems with a given condition number is closely related to computing the volume of a tubular neighborhood around a manifold, a topic studied rigorously in the mathematical discipline of *geometric probability* [80–82]. The conditioning of computational problems from the perspective of geometric probability is examined in depth in the book by Bürgisser and Cucker [83]. Additionally, estimates of the distance from a matrix to the set of matrices with multiple eigenvalues are provided in [84–88].

3.6.1. Distance to the Set of Singular Matrices

The *distance of a given matrix* $A \in \mathbb{C}^{n \times n}$ *to the variety of singular cases* \mathcal{S} is defined as

$$\text{dist}(A, \mathcal{S}) = \min \|A - Z\| : Z \in \mathcal{S} \geq 0. \quad (19)$$

To define the distance between matrices in $\mathbb{C}^{n \times n}$, we will use the 2-norm or the Frobenius norm, noting that similar results hold if another matrix norm is used.

Consider first the matrix inversion. In this setting, the following classic result holds [89], [83], Ch. 1:

Theorem 5. *Let* $A \in \mathbb{C}^{n \times n}$ *be nonsingular. Then*

$$\text{dist}(A, \mathcal{S}) = \frac{1}{\|A^{-1}\|}.$$

By defining $\text{cond}_F(A) := \infty$ for a singular matrix, we immediately obtain the relationship between the distance to singularity and the condition number.

Corollary 1. *For any nonzero* $A \in \mathbb{C}^{n \times n}$, *the following holds:*

$$\text{cond}_F(A) = \frac{\|A\|_F}{\text{dist}_F(A, \mathcal{S})}.$$

This results shows that for a normalized problem with $\|A\|_F = 1$, the condition number of the matrix in respect to inversion is inversely proportional to the distance from A to the set of singular matrices. In other words, the closer a matrix is to singularity, the larger its condition number, and hence the more sensitive it is to perturbations.

Example 13. *Figure 28 illustrates the hypersurfaces of constant condition number with respect to inversion, $\text{cond}_F(A) = 5, 10, 20$, for the matrix*

$$A = \begin{bmatrix} a & 1 \\ c & b \end{bmatrix}$$

considered in Examples 8 and 9.

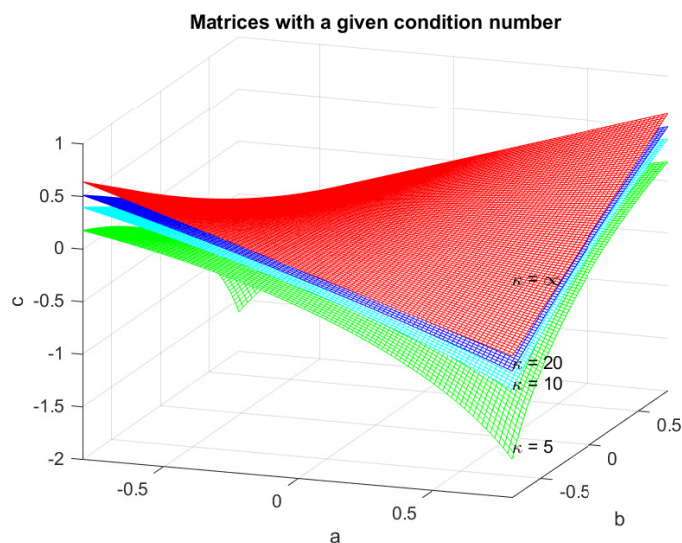


Figure 28. Third-order matrices with different condition numbers in the parameter space.

3.6.2. Distance to the Set of Defect Matrices

Consider now the eigenvalue problem. We have the following result.

Theorem 6. [78] *The distance of the matrix A to the set of matrices with multiple eigenvalues satisfies*

$$\text{dist}_F^\lambda(A, \mathcal{S}) \leq \frac{\sqrt{2}\|A\|_F}{\|P\|_F},$$

where P is the spectral projector associated with the eigenvalue of interest.

Using the expression for the eigenvalue condition number (16), and assuming that $\|A\|_F = 1$, we obtain

$$\text{dist}_F^\lambda(A, \mathcal{S}) \leq \frac{\sqrt{2}}{\kappa_\lambda^{\text{abs}}(A)}$$

which confirms that the eigenvalue condition number of the normalized problem is inversely proportional of the distance to the set of ill-posed problems.

In this way, the set of problems whose condition number is at least K is approximately the set of problems within distance C/K (with C a constant) from the variety of ill-posed matrices. As one approaches \mathcal{S} , the conditioning of the problem worsens, which is why \mathcal{S} is called *pejorative manifold* (from the Latin word *pejorare* - to make worse) by Kahan [90].

The eigenvalue problem for matrices with ill-conditioned eigenvalues is studied in [91,92].

3.7. Probabilistic Distribution of Condition Numbers

The idea of the *probabilistic distribution analysis* in the parameter space of a property characterizing the computational problem can be described in general terms [83], Ch. 2. The parameter space is endowed with a probability distribution, and a certain real-valued function $g(x)$ defined on this space is considered as a random variable. The goal is to estimate quantities such as the probability that $g(x) \geq K$ for a given K , which provides information about the behavior of g .

In this section, we show that the geometric structure of computational problems allows one to estimate the probability distribution of problems with a given condition number $\kappa(A)$ in the parameter space. To this end, we exploit the fact that the multiplication by a scalar does not change $\kappa(A)$, i.e., the condition number is homogeneous in the parameter space. This permits normalization of the problems to unit norm, so it suffices to consider only problems that lie on the unit sphere in \mathbb{R}^N or \mathbb{C}^N .

Due to the homogeneity of the condition number, its distribution in the parameter space induces the same distribution of $x/\|x\|$ over the unit sphere. It is natural to assume that problems are uniformly distributed in the parameter space, since each problem is as likely as any other. The uniformity allows us to bound the volume of the set of problems with condition number at least K , which lie within distance C/K of the variety of ill-posed problems. Consequently, the probability that $\kappa(A) > K$ is proportional to the volume of the corresponding set of problems.

Figure 29 shows an interpretation of the probabilistic distribution of the condition number in three-dimensional space. Let $T(\mathcal{S}, \epsilon)$ denote the set of all points in the unit ball $B(1)^N$ that lie within distance ϵ of \mathcal{S} ,

$$T(\mathcal{S}, \epsilon) = \{z : \|z\| \leq 1, \text{dist}(z, \mathcal{S}) \leq \epsilon\}.$$

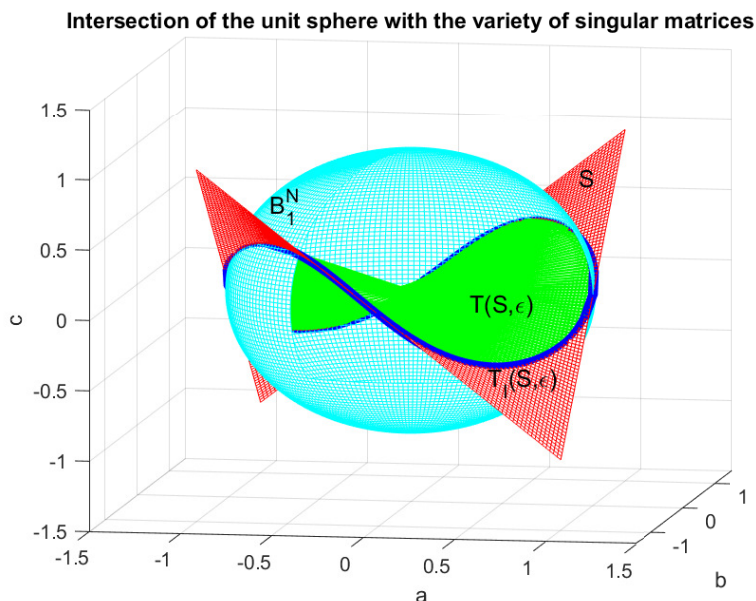


Figure 29. Intersection of the unit sphere with the variety of singular problems.

Such variety represents a tubular neighborhood (Sect. 2.7).

The ratio

$$\frac{\text{vol}(T(\mathcal{S}, \epsilon))}{\text{vol}(B_1^N)},$$

where $\text{vol}(\mathcal{M})$ denotes the volume of a manifold \mathcal{M} , gives the fraction of the unit ball within the distance ϵ of \mathcal{S} .

We are interested in the part of $T(\mathcal{S}, \epsilon)$ that intersects the unit ball, namely,

$$T_I(\mathcal{S}, \epsilon) = T(\mathcal{S}, \epsilon) \cap B_1^N.$$

Note that $T_I(\mathcal{S}, \epsilon)$ contains all singular problems lying on the unit sphere after normalization. The volume of this set is used to determine the probability distribution of the scaled singular problems $x/\|x\|$ in the ϵ -neighborhood of \mathcal{S} . By definition, this probability is given by the ratio of the volume of $T_I(\mathcal{S}, \epsilon)$ to the volume of the unit ball B_1^N , that is,

$$\text{Prob}(\text{dist}(x, \mathcal{S}) \leq \epsilon) = \frac{\text{vol}_{N-1}(T_I(\mathcal{S}, \epsilon))}{\text{vol}(B_1^N)}, \quad (20)$$

According to this definition it holds that

$$0 < \text{Prob}(\text{dist}(x, \mathcal{S}) \leq \epsilon) < 1.$$

Example 14. Consider the matrix

$$\begin{bmatrix} a & 1 \\ c & b \end{bmatrix}$$

from Examples 8 and 9. The unit ball B_1^3 of all matrices with unit Frobenius norm is given by the equation

$$a^2 + b^2 + c^2 = 1.$$

The variety of singular matrices ($\det(A) = ab - c = 0$) that lie on the unit sphere, is described by

$$b = \pm \sqrt{\frac{1-a^2}{1+a^2}}, \quad c = ab.$$

The manifold \mathcal{S} of singular matrices is represented by a hyperbolic paraboloid which is a surface with codimension 1. The sets $T(\mathcal{S}, \epsilon)$ and $T_I(\mathcal{S}, \epsilon)$ are tubular neighborhoods.

To determine $\text{Prob}(\text{dist}(x, \mathcal{S}) \leq \epsilon)$, it is necessary to compute the quantities $\text{vol}_{N-1}(T_I(\mathcal{S}, \epsilon))$ and $\text{vol}(B_1^N)$.

Proposition 2. The volume of the unit ball is given by the formula

$$\text{vol}(B_1^N) = \frac{2\pi^{n/2}}{n\Gamma(n/2)},$$

where the gamma-function $\Gamma(n)$ for a positive integer n is computed from

$$\Gamma(n) = n!.$$

Various proofs of this classical result can be found in [93].

Determining the quantity $\text{vol}_{N-1}(T_I(\mathcal{S}, \epsilon))$ is related to computing the volume of a tubular neighborhood of a real or complex manifold and represents a difficult problem. Using formulas for the volumes of tubular neighborhoods derived in [94], [76], the following theorem was proved in [78], providing an upper bound for $\text{vol}_{N-1}(T_I(\mathcal{S}, \epsilon))$ in the complex case.

Theorem 7. (Volume of a complex tubular neighborhood). [78] Assume that \mathcal{M} is a $2d$ -dimensional complex manifold in \mathbb{C}^N . Let $T_I(\mathcal{S}, \epsilon)$ be the part of the unit ball in \mathbb{C}^N that lies in a distance $\epsilon \leq 1$ from \mathcal{M} . Then

$$\text{vol}_{N-1}(T_I(\mathcal{S}, \epsilon)) \leq \frac{\sqrt{\pi}\Gamma(N+1/2)}{\Gamma(N-d+1/2)\Gamma(d+1/2)} e^2 N^2 (N-1)^{2N-2d-2} \text{deg}(\mathcal{M}) \epsilon^{2(N-d)} (1+N\epsilon)^{2d}.$$

In the above expression, $e \approx 2.7183$ denotes Euler's number, and $\text{deg}(\mathcal{M})$ is the so-called **degree** of \mathcal{M} , which generalizes the notion of the degree of a polynomial and is defined as the number of intersection points of an n -dimensional manifold \mathcal{P} in \mathbb{R}^N with an $(N-n)$ -dimensional affine subspace of \mathbb{R}^N . If \mathcal{M} is a hypersurface ($d = N-1$), this upper bound can be improved to

$$\text{vol}_{N-1}(T_I(\mathcal{S}, \epsilon)) \leq e^2 N^3 \text{deg}(\mathcal{M}) \epsilon^2 (1+N\epsilon)^{2(N-1)}.$$

The expressions for the volume of a real tubular neighborhood are more complicated and can be found in [95], [96], [83], Ch. 21.

3.7.1. Probability Conditioning of Matrix Inversion

For computational convenience, when determining the probability of occurrence of a matrix with a given condition number, instead of the usual condition number $\kappa_2(A) = \|A\|_2 \|A^{-1}\|_2$, we shall use the nearly equivalent scaled condition number $\kappa(A) = \|A\|_F \|A^{-1}\|_2$. Since $\|A\|_2 \leq \|A\|_F$, it follows that $\kappa_2(A) \leq \kappa(A)$.

Using Theorem 7, the following result concerning the probabilistic distribution of the matrix inversion problem was proved in [78].

Theorem 8. Let $A \in \mathbb{C}^{n \times n}$ be a random matrix distributed in the parameter space such that $A/\|A\|_F$ is uniformly distributed on the unit sphere. Define

$$\kappa(A) = \|A\|_F \|A^{-1}\|_2 = \|A^{-1}\|_2 = 1/\text{dist}_2(A, \mathcal{S}),$$

where \mathcal{S} denotes the set of singular matrices and $\text{dist}_2(A, \mathcal{S}) = 1/\|A^{-1}\|_2$. Then for a given number $K > 0$, the probability that $\kappa(A) \geq K$ satisfies

$$\frac{(1 - 1/K)^{2n^2-2}}{2n^4K^2} \leq \text{Prob}(\kappa(A) \geq K) \leq \frac{e^2n^5(1 + n^2/K)^{2n^2-2}}{K^2} \quad (21)$$

and

$$\text{Prob}(\kappa(A) \geq K) = \frac{n^2(n-1)}{K^2} + o\left(\frac{1}{K^2}\right). \quad (22)$$

Expression (21) provides lower and upper bounds for the probability, while (22) gives its asymptotic behavior as $K \rightarrow \infty$.

Theorem 8 was refined in [97], where the following exact result was obtained:

Theorem 9. For a complex $n \times n$ matrix A , the probability that the condition number $\kappa(A)$ satisfies $\kappa(A) \geq K$ is

$$\text{Prob}(\kappa(A) \geq K) = 1 - (1 - n/K^2)^{n^2-1}, \quad K > \sqrt{n}. \quad (23)$$

For large K and $n > 20$, the asymptotic estimate holds:

$$\text{Prob}(\kappa(A) \geq K) \approx n^{3/2}/K, \quad K \gg \sqrt{n}, \quad n \gg 1.$$

According to this result, the probability that the condition number of a matrix exceeds a given value K is inversely proportional to K . That is, as a matrix approaches the set of singular matrices, the condition number

$$\kappa(A) = \frac{\|A\|_F}{\text{dist}_2(A, \mathcal{S})}$$

increases, but the likelihood of encountering a very ill-conditioned matrix decreases.

A graphical interpretation of equation (23) for $n = 10, 100, 1000$ is shown in Figure 30, where the probability estimate is plotted as a function of $\text{dist}(A, \mathcal{S})/\|A\|_F = 1/K$. As the matrix dimension increases, the probability of encountering a larger condition number also increases.

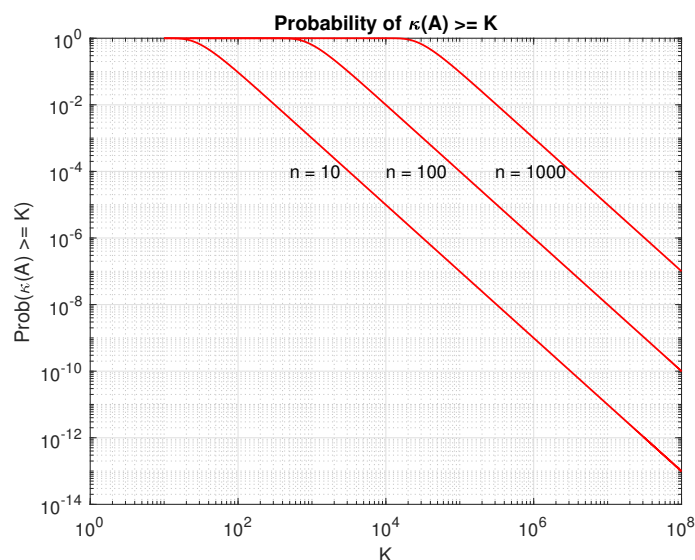


Figure 30. Probability distribution of the condition number with respect to inversion.

Using a similar technique, the following result concerning the condition number with respect to inversion was obtained in [98]:

Theorem 10. For all $n \geq 1$, $0 < \varepsilon \leq 1$, it holds that

$$\sup_{Z \in \mathbb{S}^{n^2-1}} \mathbf{E}_{Z \in B_\varepsilon(A)} (\ln \kappa(A)(Z)) \leq 6 \ln n + 2 \ln \frac{1}{\varepsilon} + 4.7,$$

where the expectation \mathbf{E} is taken over all Z uniformly distributed in the open ball $B_\varepsilon(A)$ of radius ε centered at A on the unit sphere \mathbb{S}^{n^2-1} .

This result reflects the fact that, as $\varepsilon \rightarrow 0$, the condition number tends to infinity.

3.7.2. Probability Conditioning of Eigenvalues

A result similar to Theorem 8, can be obtained with respect to the eigenvalue problem.

Theorem 11. [78] Let $A \in \mathbb{C}^{n \times n}$ be a random matrix distributed in the parameter space such that $A/\|A\|_F$ is uniformly distributed on the unit sphere. Define

$$\kappa_\lambda(A) := \max_{\lambda(A)} \|P_{\lambda(A)}\|_F = \sqrt{2}/\text{dist}_F \lambda(A, \mathcal{S}),$$

where the maximum is taken over all eigenvalues $\lambda(A)$ of A , and $P_{\lambda(A)}$ denotes the spectral projector associated with $\lambda(A)$. Then for any given number $K > 0$ the probability that $\kappa_\lambda(A) \geq K$, satisfies

$$\text{Prob}(\kappa_\lambda(A) \geq K) \leq \frac{e^2 n^5 (n-1) (1 + \sqrt{2} n^2 / K)^{2n^2-2}}{K^2}. \quad (24)$$

In Figure 31 we show the upper probability bound given by (24) for $n = 5, 10, 15$. This bound is meaningful for $K > 11828$, which means that for $\text{dist}(A, \mathcal{S}) > 1.1956 \times 10^{-4}$ the matrices with $\kappa_\lambda(A) \leq 11828$ appear with a probability equal to 1.

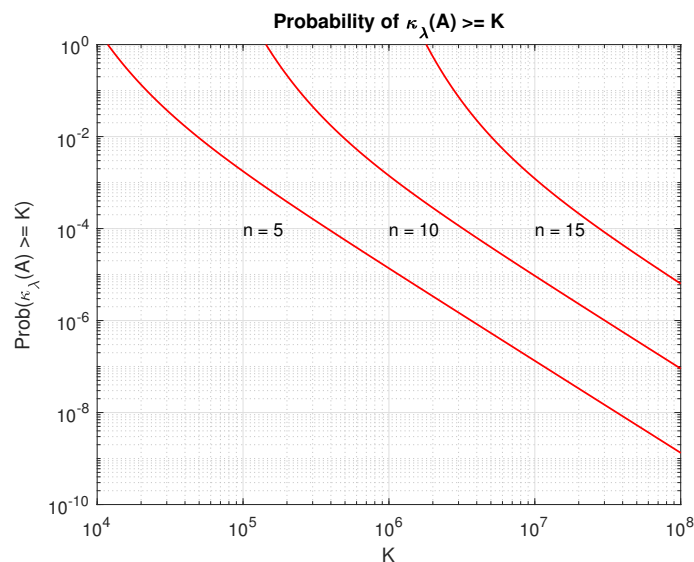


Figure 31. Probability distribution of the eigenvalue condition number.

An analogue of Theorem 10 is given by the following result.

Theorem 12. [98] For all $n \geq 1$ and $0 < \varepsilon \leq 1$, the following holds:

(a) For all real matrices,

$$\sup_{A \in \mathbb{S}^{n^2-1}} \mathbf{E}_{Z \in B_\varepsilon(A)} (\ln \kappa_\lambda(Z)) \leq 8 \ln n + 2 \ln \frac{1}{\varepsilon} + 5.1.$$

(b) For all complex matrices,

$$\sup_{A \in \mathbb{S}^{2n^2-1}} \mathbf{E}_{Z \in B_\varepsilon(A)} (\ln \kappa_\lambda(Z)) \leq 8 \ln n + 2 \ln \frac{1}{\varepsilon} + 6.5.$$

Similar results are obtained for the polynomial zero finding, see [78].

The assumed uniform distribution is a continuous model which is a good approximation only as long as the finite-precision numbers of the computer arithmetic are dense enough to resemble the continuum. For a detailed discussion of this limitation of the method, see [78].

4. Geometry of Matrix Rank

The variety of singular cases in computational problems can have a highly complex structure, depending on the problem being solved. In this section, we focus on the geometric structure of the variety of singular cases in matrix space, particularly those associated with the rank of rectangular matrices.

4.1. Orbits of Matrices with Constant Rank

In studying the manifolds of rectangular matrices, different matrix manifolds can be obtained through equivalent transformations of matrices with a fixed rank. Specifically, in the space of $m \times n$ rectangular matrices, the set of all matrices equivalent to a given matrix A form a smooth manifold in $\mathbb{C}^{m \times n}$. This manifold is defined as

$$\mathcal{O}(A) = \{PAQ^{-1} : \det(P)\det(Q) \neq 0\}, \quad (25)$$

and is called the *orbit* of A . Since the equivalent transformations preserve the rank of a matrix, each orbit consists of all matrices of a fixed rank r and the entire space $\mathbb{C}^{m \times n}$ is partitioned into orbits containing matrices of the same rank.

Consider the linear operator

$$\mathcal{A} : \mathbb{C}^n \rightarrow \mathbb{C}^m$$

represented by the rectangular matrix $A \in \mathbb{C}^{m \times n}$. The finite-dimensional space of all linear operators from \mathbb{C}^n to \mathbb{C}^m is isomorphic to the vector space $\mathbb{C}^{m \cdot n}$. The linear operators \mathcal{A} of maximal rank $r = \min\{m, n\}$ form an everywhere dense subset of $\mathbb{C}^{m \cdot n}$. Such operators are called *regular* or *non-singular*. Hence, the non-singularity is a generic property of the operators in $\mathbb{C}^{m \cdot n}$. According to a well known results from linear algebra, a non-singular operator can be represented in a suitable choice of bases by the $m \times n$ matrix

$$A_0 = \begin{bmatrix} I_r \\ 0_{m-r,r} \end{bmatrix} \text{ if } m \geq r = n$$

or

$$A_0 = \begin{bmatrix} I_r & 0_{r,n-r} \end{bmatrix} \text{ if } r = m \leq n.$$

If $r < \min\{m, n\}$, the operator is *singular* and, in a suitable choice of bases, it can be represented by the $m \times n$ matrix

$$A_0 = \begin{bmatrix} I_r & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{bmatrix}.$$

The matrix A_0 is called the *normal form* of A and the differences $m - r$ and $n - r$ are referred to as *coranks*.

4.2. Dimension of an Orbit with Fixed Rank

In what follows, we will show that the singular operators, corresponding to matrices A of rank $r < \min\{m, n\}$ form an orbit in $\mathbb{C}^{m \times n}$, whose codimension is the product $(m - r)(n - r)$.

For a rectangular matrix $A \in \mathbb{C}^{m \times n}$, define the linear operator

$$\mathcal{L}_A : (X, Y) \rightarrow XA - AY, \quad (X, Y) \in \mathbb{C}^{m \times m} \times \mathbb{C}^{n \times n}.$$

Let $T_A \mathcal{O}(A)$ denote the tangent space to the orbit $\mathcal{O}(A)$ at A and let $\mathcal{R}(\mathcal{L}_A)$ denote the image of \mathcal{L}_A . Then we have the following result.

Lemma 1.

$$T_A \mathcal{O}(A) = \mathcal{R}(\mathcal{L}_A).$$

Proof. Consider the nonlinear transformation ψ from the sets of $m \times m$ and $n \times n$ nonsingular matrices into the space $\mathbb{C}^{m \times n}$, defined by

$$\psi(P, A, Q) = PAQ^{-1}.$$

Assume that the matrices $|X|$ and $|Y|$ are sufficiently small, and set

$$P = I_m + X, \quad Q = I_n + Y,$$

so that P and Q remain nonsingular. Then, for $|X| \rightarrow 0$ and $|Y| \rightarrow 0$, we have the expansions $P^{-1} = I_m - X + O(|X|^2)$, $Q^{-1} = I_n - Y + O(|Y|^2)$. Substituting these into ψ gives

$$\begin{aligned} \psi(I_m + X, A, I_n + Y) &= (I_m + X)A(I_n + Y)^{-1} \\ &= A + XA - AY + O(|Y|^2) \\ &= \psi(I_m, A, I_n) + \mathcal{L}_A(X, Y) + O(|Y|^2), \end{aligned} \tag{26}$$

where $\mathcal{L}_A(X, Y) = XA - AY$. Hence, the tangent vectors to the orbit $\mathcal{O}(A)$ are given by the differential

$$D\psi(X, Y) = \mathcal{L}_A(X, Y) = XA - AY.$$

Evaluating the differential at the identity, $P = I_m$, $Q = I_n$, we obtain

$$T_A \mathcal{O}(A) = \mathcal{R}(\mathcal{L}_A),$$

i.e., the tangent space to the orbit at A is exactly the range of the linear operator \mathcal{L}_A .

The dimension of the orbit of matrices with fixed rank can be determined by noting that

$$\dim(\mathcal{O}(A)) = T_A \mathcal{O}(A),$$

where $T_A \mathcal{O}(A)$ is the tangent space at A . Define the subspace

$$Z_A = \{(X, Y) \in \mathbb{C}^{m \times m} \times \mathbb{C}^{n \times n} : XA - AY = 0\}.$$

Clearly, Z_A is the null space of the linear operator \mathcal{L}_A . Since for any linear operator, the dimension of the domain equals the sum of the dimensions of the range and null space, we have

$$\mathbb{C}^{m \times n} = \mathcal{R}(\mathcal{L}_A) + \mathcal{N}(\mathcal{L}_A),$$

where $\mathcal{N}(\mathcal{L}_A)$ is the null space. Therefore,

$$\begin{aligned} \dim(\mathbb{C}^{m \times n}) &= \dim(\mathcal{R}(\mathcal{L}_A)) + \dim(\mathcal{N}(\mathcal{L}_A)) \\ &= \dim(T_A \mathcal{O}) + \dim(Z_A) \\ &= \dim(\mathcal{O}(A)) + \dim(Z_A). \end{aligned}$$

This shows that the dimension of the orbit is equal to the dimension of the ambient space minus the dimension of the null space of \mathcal{L}_A .

Thus, we arrive at the important relationship

$$\dim(Z_A) = \dim(\mathbb{C}^{m \times n}) - \dim(\mathcal{O}(A)) = \text{codim}(\mathcal{O}(A)), \quad (27)$$

i.e., the dimension of the subspace Z_A is equal to the dimension of the normal space $N_{\mathcal{O}}$, which in turn equals the codimension of the orbit $\mathcal{O}(A)$. Since $\dim(Z_A)$ is invariant under equivalent transformations of A , it follows that

$$\dim(Z_A) = \dim(Z_{A_0}),$$

where $Z_{A_0} = \{(X, Y) : XA_0 - A_0Y = 0\}$ and

$$A_0 = \begin{bmatrix} I_r & 0_{r, n-r} \\ 0_{m-r, r} & 0_{m-r, n-r} \end{bmatrix}.$$

is the normal form of a rank- r matrix.

Partitioning the matrices X and Y as

$$X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}, X_{11} \in \mathbb{C}^{r \times r}, Y = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix}, Y_{11} \in \mathbb{C}^{r \times r},$$

we have

$$\mathcal{L}_A(X, Y) = \begin{bmatrix} X_{11} & 0_{r, n-r} \\ X_{21} & 0_{m-r, n-r} \end{bmatrix} - \begin{bmatrix} Y_{11} & Y_{12} \\ 0_{m-r, r} & 0_{m-r, n-r} \end{bmatrix} = \begin{bmatrix} X_{11} - Y_{11} & -Y_{12} \\ X_{21} & 0_{m-r, n-r} \end{bmatrix}.$$

From this expression, it follows that

$$\begin{aligned} \dim(\mathcal{R})(\mathcal{L}_A) &= \dim(\mathbb{C}^{r \times r} \times \mathbb{C}^{(m-r) \times r} \times \mathbb{C}^{r \times (n-r)}) \\ &= \dim(\mathbb{C}^{r \times r}) + \dim(\mathbb{C}^{(m-r) \times r}) + \dim(\mathbb{C}^{r \times (n-r)}) \\ &= r^2 + (m-r)r + (n-r)r. \end{aligned}$$

Hence, the dimension of the null space of \mathcal{L}_A is

$$\dim(Z_A) = \dim(\mathcal{P}(\mathcal{L}_A)) = mn - r^2 - (m-r)r - (n-r)r = (m-r)(n-r)$$

which leads to the following important result.

Theorem 13. *The codimension of rank- r matrices in $\mathbb{C}^{m \times n}$ is $(m-r)(n-r)$.*

The dimension and codimension counts are illustrated in Figure 32. The dimension of the orbit is given by

$$\dim(\mathcal{O}(A)) = mn - (m-r)(n-r) = (m+n-r)r$$

The perturbation with the minimum number of parameters that acts in the normal space of $\mathcal{O}(A_0)$ has the form

$$\delta A = \left[\begin{array}{c|cccc} 0_{r \times r} & & & & \\ \hline & \delta_{11} & \delta_{12} & \dots & \delta_{1,n-r} \\ 0_{(m-r) \times r} & \delta_{21} & \delta_{22} & \dots & \delta_{2,n-r} \\ & \vdots & \vdots & \ddots & \vdots \\ & \delta_{m-r,1} & \delta_{m-r,2} & \dots & \delta_{m-r,n-r} \end{array} \right],$$

where δ is an $(m - r) \times (n - r)$ block with independent entries δ_{ij} . This perturbation lies in the normal space to the orbit $\mathcal{O}(A_0)$ and is transversal (i.e., in general position) to $\mathcal{O}(A_0)$.

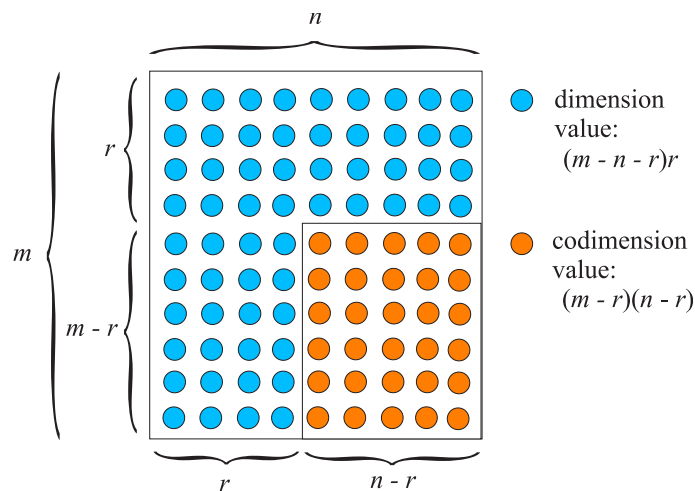


Figure 32. Dimension and codimension of the orbit $\mathcal{O}(A)$.

The concept of the matrix rank was introduced by the German mathematician Ferdinand Georg Frobenius in 1879 [99], albeit implicitly in connection with determinants. Theorem 13 has been proved using various methods in ([83], p. 470), and [100,101].

4.3. Stratification of Orbits with Fixed Rank

The partition of a space into a finite number of submanifolds defined by algebraic equations and inequalities is called a **stratification**. The stratification means that the matrix space $\mathbb{C}^{m \times n}$ is decomposed into manifolds, called *strata*, which are arranged in different layers. The partition of the matrix space into orbits with fixed rank is an example of *matrix stratification*.

We say that a complex manifold \mathcal{M} is *embedded* in another complex manifold \mathcal{P} , if \mathcal{M} is contained in the closure $\overline{\mathcal{P}}$ of \mathcal{P} , which we denote by

$$\mathcal{M} \subseteq \overline{\mathcal{P}}.$$

Every orbit is embedded in itself and in all orbits of equal or higher rank (equivalently, equal or lower codimension).

Theorem 14. (Orbit Embedding Theorem). *In the matrix space $\mathbb{C}^{m \times n}$, the orbit \mathcal{O}_r of matrices of rank r is embedded in the orbit \mathcal{O}_s if and only if $r \leq s$ or, equivalently, if and only if*

$$\text{codim}(\mathcal{O}_r) \geq \text{codim}(\mathcal{O}_s).$$

More precisely, the following chain of inclusions holds:

$$\mathcal{O}_0 \subseteq \mathcal{O}_1 \subseteq \dots \subseteq \mathcal{O}_{\min\{m,n\}}.$$

Theorem 14 describes the stratification of the matrix space $\mathbb{C}^{m \times n}$ into orbits of different ranks. According to this result, if a matrix A belongs to the orbit \mathcal{O}_r , it also lies in the closure of all orbits \mathcal{O}_s with higher rank $s \geq r$. These higher-rank orbits have strictly smaller codimension than the orbit containing A .

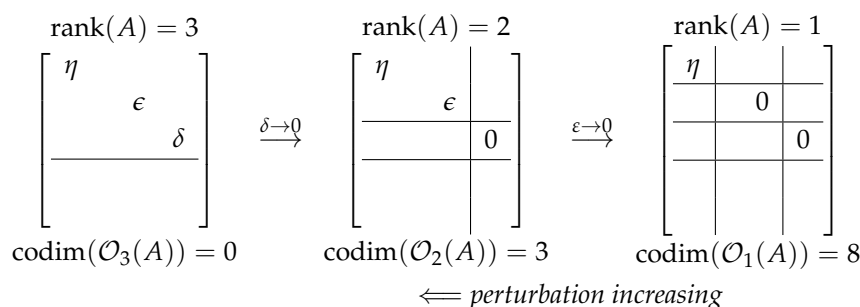
Example 15. Consider families of 6×4 matrices of different rank. For each rank $r = 0, 1, 2, 3, 4$ there exists an orbit \mathcal{O}_r consisting of all 6×4 matrices of rank r . The codimension of each orbit is given by $(m - r)(n - r)$. These orbits form a stratification of the matrix space $\mathbb{C}^{6 \times 4}$, which can be summarized as follows:

rank :	0	<	1	<	2	<	3	<	4
orbit :	\mathcal{O}_0	\subseteq	\mathcal{O}_1	\subseteq	\mathcal{O}_2	\subseteq	\mathcal{O}_3	\subseteq	\mathcal{O}_4
dimension :	0	<	9	<	16	<	21	<	24
codimension :	24	>	15	>	8	>	3	>	0

The table clearly illustrates the inclusion relations between the orbits: as the rank increases, the dimension of the orbit increases while its codimension decreases. The full-rank orbit \mathcal{O}_4 is open and dense in $\mathbb{C}^{6 \times 4}$, whereas the lower-rank orbits form boundary strata of increasing codimension.

The next example illustrates how infinitely small perturbations can cause a matrix A of a given rank to move from an orbit of higher codimension to an orbit of lower codimension.

Example 16. Consider rank transitions of a 5×3 matrix (blank entries are understood to be 0):



Note that perturbations of general position act from right to left increasing the rank of the matrix. In contrast, decreasing the rank from left to right requires perturbations with a special (diagonal) structure. Such perturbations lie in the tangent space of the corresponding matrix orbit $\mathcal{O}(A)$ and are therefore not in general position with respect to $\mathcal{O}(A)$ and not generic.

From the figure we see that the rightmost 5×3 matrix A with rank(A) = 1 (and codim($\mathcal{O}_1(A)$) = (5 - 1)(3 - 1) = 8) is arbitrarily close to matrices with rank(A) = 2 (whose orbit has codimension codim($\mathcal{O}_2(A)$) = 3). In turn, these matrices are arbitrary close to matrices with rank(A) = 3 (for which codim($\mathcal{O}_3(A)$) = 0).

Thus, by applying a sufficiently small perturbation that is in general position to A , one can move from the orbit $\mathcal{O}_1(A)$, rank(A) = 1, to the orbit $\mathcal{O}_2(A)$, rank(A) = 2. In other words, the orbit $\mathcal{O}_1(A)$, rank(A) = 1 is contained in the closure of $\mathcal{O}_2(A)$, and $\mathcal{O}_2(A)$, is said to cover $\mathcal{O}_1(A)$. Formally, this relation is written as

$$\mathcal{O}_1(A) \subset \overline{\mathcal{O}_2},$$

where $\overline{\mathcal{O}_2}$ denotes the closure of the orbit of rank-2 matrices.

Similarly, a small perturbation in general position allows one to move from $\mathcal{O}_2(A)$, rank(A) = 2 to $\mathcal{O}_3(A)$, rank(A) = 3. This implies that

$$\mathcal{O}_2(A) \subset \overline{\mathcal{O}_3}.$$

Taken together, these closure relations show that, via a sufficiently small perturbation, one can move directly from $\mathcal{O}_1(A)$, $\text{rank}(A) = 1$ to $\mathcal{O}_3(A)$, $\text{rank}(A) = 3$, passing through intermediate orbits of lower codimension.

It is important in numerical computations that adding a perturbation in general position to a matrix A can only decrease the codimension of the orbit in which the perturbed matrix lies; it can never increase it. This implies that, due to rounding errors, numerical algorithms effectively operate on matrices that belong to orbits (manifolds) of codimension 0, i.e., the full-rank or maximal-rank orbits.

For orbits of matrices with fixed rank, it is possible to compute exactly the distance in the matrix space from a given matrix to an orbit corresponding to a prescribed rank.

Theorem 15. (Schmidt-Mirsky Theorem). [45], Ch. 1 Let $A \in \mathbb{C}^{m \times n}$ be a matrix with singular values

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0,$$

where $r = \text{rank}(A)$. Then the distance from A to an orbit of matrices with rank $k < r$ is given by

$$\begin{aligned} \text{dist}_2(A, \mathcal{O}(B : \text{rank}(B) = k)) &= \sigma_{k+1}, \quad k = 1, 2, \dots, r-1, \\ \text{dist}_F(A, \mathcal{O}(B : \text{rank}(B) = k)) &= \sqrt{\sigma_{k+1}^2 + \dots + \sigma_r^2}, \quad k = 1, 2, \dots, r-1. \end{aligned}$$

Thus, the matrix A lies at distances $\sigma_2, \sigma_3, \dots, \sigma_r$ from the orbits corresponding to rank $1, 2, \dots, r-1$, respectively.

4.4. Numerical Rank of a Matrix

An important task arising in matrix computations is determining the rank of a given matrix A in the presence of uncertainties in its entries. This problem is closely related to solving linear systems of less-than-full rank and to the numerical determination of the Jordan structure of a matrix. The difficulty in determining the rank stems from the fact that it is not a continuous function of the matrix entries and may change abruptly under arbitrarily small perturbations of these entries.

Let $A \in \mathbb{C}^{m \times n}$, and without loss of generality assume that $m \geq n$. Let

$$A = U\Sigma V^H$$

be the singular value decomposition (SVD) of A , where $U \in \mathbb{C}^{m \times m}$, $V \in \mathbb{C}^{n \times n}$ are unitary matrices, and

$$\Sigma = \text{diag}(\sigma_1, \dots, \sigma_k, \dots, \sigma_n)$$

contains the singular values of A , with $\sigma_1 \geq \dots \geq \sigma_k \geq \dots \geq \sigma_n$. Note that $\sigma_1 = \|A\|_2$.

The “theoretical” rank of A is defined as the number of nonzero singular values. In practice, the matrix A contains measurement, approximation, and discretization errors. Therefore, instead of the exact rank of A , we determine the rank of a perturbed matrix $A + \delta A$, where the perturbation δA satisfies

$$\|\delta A\|_2 \leq \text{tol}\|A\|_2$$

for some small positive number tol . The quantity tol can be interpreted as the relative uncertainty in A .

These considerations lead to the concept of the *numerical rank of a matrix* relative to the tolerance $\text{tol}\|A\|_2$, defined as

$$k = k(A, \text{tol}) := \min_{\|\delta A\|_2 \leq \text{tol}\|A\|_2} \text{rank}(A + \delta A). \quad (28)$$

Unlike the “theoretical”; rank, the numerical rank is stable in the sense that perturbations smaller than the tolerance $\text{tol}\|A\|_2$ will not change the rank of A .

The concept of the numerical rank of a matrix is discussed in detail in [46], Sec. 5.4, [45], Ch.], [53], Sec. 3.5, and [102], Ch. 3; see also [103–105]. Efficient procedures for determining the numerical rank are described in [106,107] and [108].

In connection with the stability of the numerical rank, the following can be noted.

In general terms, a mathematical object is called *structurally stable*, if its structure remains unchanged under perturbations of the object's parameters. If the object remains stable under large parameter variations, it is called *robust*.

The normal canonical form of a rectangular matrix is an example of an object that is structurally unstable. The concepts of structural stability, genericity, and transversality, when applied to dynamical system, are discussed in depth in [109], Ch. 16, [110], Ch. 4, [111], Ch. 12, and [112], Ch. 3.

5. Geometry of Jordan form

5.1. Orbits of Matrices with Fixed Jordan Form

In the space of $n \times n$ square matrices, the set of all matrices similar to a given matrix A form a manifold in $\mathbb{C}^{n \times n}$. This manifold, defined by

$$\mathcal{O}(A) = \{PAP^{-1} : \det(P) \neq 0\}, \quad (29)$$

is called the *orbit* of the matrix A . All matrices lying in the same orbit have the same eigenvalues and the same dimensions of the Jordan blocks.

The *bundle* $\mathcal{B}(A)$ of the matrix A is defined as the union of all orbits. It consists of all matrices whose Jordan canonical forms differ only in their eigenvalues while having the same number of distinct eigenvalues and the same sizes of the corresponding Jordan blocks (i.e., the same Segre characteristics). If two matrices have identical Jordan structures but different distinct eigenvalues, they belong to the same bundle. For example, all diagonal matrices with simple eigenvalues form a single bundle.

Each bundle is a manifold in the space of matrices, whose strata are the individual orbits. Within a given bundle, an orbit consists precisely of those matrices that share the same eigenvalues.

As an illustration, consider the two matrices with Jordan forms

$$J_1 = \left[\begin{array}{cc|cc} -1 & 1 & & \\ & -1 & & \\ \hline & & 1 & \\ \hline & & & 2 \end{array} \right], \quad J_2 = \left[\begin{array}{cc|cc} 2 & 1 & & \\ & 2 & & \\ \hline & & 3 & \\ \hline & & & 4 \end{array} \right].$$

These matrices lie in different orbits, since their eigenvalues differ, but they belong to the same bundle because their Jordan block structures are identical.

The most important characteristic of orbits and bundles are their dimensions. The dimension of an orbit, denoted by $\dim(\mathcal{O}(A))$, is equal to the dimension of its tangent space $T_{\mathcal{O}}(x)$. In practice, it is often more convenient to work with the *codimension* of $\mathcal{O}(A)$ denoted by $\text{codim}(\mathcal{O}(A))$, which is equal to the dimension of the normal space $N_{\mathcal{O}}(x)$. Since

$$T_{\mathcal{O}}(x) + N_{\mathcal{O}}(x) = \mathbb{C}^{n \times n},$$

it follows that

$$\dim(\mathcal{O}(A)) + \text{codim}(\mathcal{O}(A)) = n^2.$$

The following result can now be established.

Theorem 16. [100] *Let*

$$\mathcal{S}_i = \{s_{1i} \geq s_{2i} \geq \dots \geq s_{q_i,i}\}$$

be the Segre characteristic of A associated with the eigenvalue λ_i . Then the codimension of the orbit $\mathcal{O}(A)$ is given by

$$\begin{aligned}\text{codim}(\mathcal{O}(A)) &= \sum_{i=1}^p \sum_{j=1}^{q_i} (2j-1)s_{ji} \\ &= \sum_{i=1}^p (s_{1i} + 3s_{2i} + 5s_{3i} + \cdots + (2q_i - 1)s_{q_i,i}),\end{aligned}\quad (30)$$

where p denotes the number of the distinct eigenvalues of A . Note that the complex conjugate eigenvalues are counted as two distinct eigenvalues.

The difference between orbits and bundles is that the eigenvalues of the matrices belonging to a bundle are not fixed. In other words, while an orbit consists of matrices similar to A with the same eigenvalues, a bundle allows the eigenvalues to vary, provided the Jordan structure remains unchanged. As a consequence, the tangent space of a bundle contains one additional dimension for each distinct eigenvalue compared to the tangent space of the corresponding orbit. Hence the codimension of the bundle $\mathcal{B}(A)$ is given by

$$\begin{aligned}\text{codim}(\mathcal{B}(A)) &= \sum_{i=1}^p (-1 + \sum_{j=1}^{q_i} (2j-1)s_{ji}) \\ &= \sum_{i=1}^p (-1 + s_{1i} + 3s_{2i} + 5s_{3i} + \cdots + (2q_i - 1)s_{q_i,i}).\end{aligned}\quad (31)$$

Comparing (30) and (31), we obtain

$$\text{codim}(\mathcal{B}(A)) = \text{codim}(\mathcal{O}(A)) - p,$$

that is, the codimension of a bundle is equal to the codimension of the corresponding orbit minus the number of distinct eigenvalues. This relation reflects the fact that a bundle possesses additional degrees of freedom – one for each distinct eigenvalue – which are absent in a fixed orbit.

Note that simple eigenvalues contribute nothing to the sum in (31). Furthermore, the codimension of a bundle does not depend on the order n of the matrix, but only on the sizes of the Jordan blocks corresponding to multiple eigenvalues. Thus, the bundle codimension provides a measure of the dependencies in the space of matrices imposed by the Jordan structure of the matrix.

The partition of the matrix space into orbits and bundles, represented by manifolds with corresponding codimensions, was introduced by the Russian mathematician Vladimir Arnold in [100] in connection with the determination of normal forms of matrices depending on a minimal number of perturbation parameters. This approach makes it possible to apply methods of differential topology to the study of matrix problems [113–115,117,118], [116], Ch. 14.

5.2. Generic and Nongeneric Jordan Bundles

Let $\mathcal{X} = \mathbb{R}^n$ or $\mathcal{X} = \mathbb{C}^n$, and denote by $\mathcal{L}\mathcal{X}$ the set of $n \times n$ matrices, or equivalently, the set of linear maps on \mathcal{X} . By writing all entries of a matrix columnwise as a vector, we may identify $\mathcal{L}\mathcal{X}$ with $\mathbb{R}^{n \cdot n}$ or $\mathbb{C}^{n \cdot n}$, respectively. Then the following important result holds [120], Sect. 5.6.

Theorem 17. *The set \mathcal{M} of matrices in $\mathcal{L}\mathcal{X}$ that have n distinct eigenvalues is open and dense in $\mathcal{L}\mathcal{X}$.*

According to this theorem, having all distinct eigenvalues is a generic property of $n \times n$ matrices, that is, “almost all” matrices have distinct eigenvalues and are therefore diagonalizable. Matrices with non-diagonal (Jordan) forms, whose Jordan blocks have specified sizes (the Segre characteristics), lie on a bundle of \mathcal{X} , whose dimension is determined by the sizes of the blocks.

In the case of two distinct eigenvalues, μ_1 and μ_2 , each orbit is determined by a fixed combination of these eigenvalues and corresponds to a plane in the 3-dimensional space. Note that the distance between planes corresponding to infinitesimally close distinct eigenvalues is itself infinitesimal. The set of matrices with simple eigenvalues forms a bundle with dimension 4, which is dense set in the space of 2×2 matrices. Such matrices are most probable, in the sense that a matrix with randomly chosen entries almost surely has distinct (simple) eigenvalues. For this reason, this case is referred to as the *most generic case*.

In Table 1 we show the different Jordan forms corresponding to the case of a single eigenvalue λ with algebraic multiplicity n . Note that the single n th order Jordan block associated with such eigenvalue represents the most generic Jordan structure in the case of multiple eigenvalue. The corresponding bundle contains all matrices A which are similar to n th-order companion matrices.

Table 1. Jordan forms corresponding to an n -tuple eigenvalue.

Most generic Jordan form

Single n th order Jordan block

$$J = \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \lambda & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda \end{bmatrix},$$

$$\begin{aligned} \mathcal{S} &= \{n\} \\ \dim(\mathcal{O}) &= n^2 - n \\ \text{codim}(\mathcal{O}) &= n \\ \dim(\mathcal{B}) &= n^2 - n + 1 \\ \text{codim}(\mathcal{B}) &= n - 1 \end{aligned}$$

Intermediate cases

$1 < q < n$ Jordan blocks
with the same eigenvalue

$$J = \begin{bmatrix} J_{s_1}(\lambda) & & & & \\ & J_{s_2}(\lambda) & & & \\ & & J_{s_3}(\lambda) & & \\ & & & \ddots & \\ & & & & J_{s_q}(\lambda) \end{bmatrix},$$

$$\begin{aligned} \mathcal{S} &= \{s_1, s_2, \dots, s_q\} \\ s_1 &\geq s_2 \geq \dots \geq s_q \geq 1 \\ n - 1 &\leq \text{codim}(\mathcal{B}) \leq n^2 - 1 \\ \text{codim}(\mathcal{B}) &= s_1 + 3s_2 + \dots + (2q - 1)s_q \end{aligned}$$

Most degenerate Jordan form

n scalar blocks with the same eigenvalue

$$J = \begin{bmatrix} \lambda & & & & \\ & \lambda & & & \\ & & \lambda & & \\ & & & \ddots & \\ & & & & \lambda \end{bmatrix},$$

$$\begin{aligned} \mathcal{S} &= \{1, 1, \dots, 1\} \\ \dim(\mathcal{O}) &= 0 \\ \text{codim}(\mathcal{O}) &= n^2 \\ \dim(\mathcal{B}) &= 1 \\ \text{codim}(\mathcal{B}) &= n^2 - 1 \end{aligned}$$

Table 2 summarizes the most generic and the most degenerate cases of $n \times n$ matrices in terms of their Jordan structures. The most generic case corresponds to matrices with n distinct eigenvalues, each forming a separate Jordan block; in this case, the orbit has maximal dimension $n^2 - n$ and codimension n , while the corresponding bundle has dimension n^2 and codimension 0. Conversely, the most degenerate case corresponds to a single eigenvalue with n scalar Jordan blocks; here, the orbit is zero-dimensional with maximal codimension n^2 , and the bundle has minimal dimension 1 and

codimension $n^2 - 1$. This table illustrates how the Jordan structure directly influences the dimensions of orbits and bundles.

Table 2. Most generic and most degenerate orbits and bundles of $n \times n$ matrices

Case	Number of Jordan blocks	Number of distinct eigenvalues	Segre characteristics	Dimensions of orbits and bundles
Most generic	n	n	$\mathcal{S} = [\{1\}, \{1\}, \dots, \{1\}]$	$\dim(\mathcal{O}) = n^2 - n$ $\text{codim}(\mathcal{O}) = n$ $\dim(\mathcal{B}) = n^2$ $\text{codim}(\mathcal{B}) = 0$
Most degenerate	n	1	$\mathcal{S} = \{1, 1, \dots, 1\}$	$\dim(\mathcal{O}) = 0$ $\text{codim}(\mathcal{O}) = n^2$ $\dim(\mathcal{B}) = 1$ $\text{codim}(\mathcal{B}) = n^2 - 1$

In the general case, when a matrix A belongs to a given bundle, it may also lie in the closure of many other bundles corresponding to different Segre characteristics. These bundles have smaller codimension than the bundle containing the original matrix, forming a hierarchy of stratification of Jordan structures. Adding a perturbation in general position decreases the codimension of the bundle into which the perturbed matrix moves. The stratification of Jordan structures has been studied by Edelman, Elmroth, and Kågström [115], who show that Jordan and Kronecker canonical forms can be represented as integer partitions. These partitions reveal closure relations of orbits and bundles through simple combinatorial rules, which can also be used to determine whether one structure is more generic than another.

5.3. The Reduction into Jordan Form as an Ill-Posed Problem

Determining the Jordan canonical form of a square nonsymmetric matrix A with defective eigenvalues by means of a computer is one of the most challenging problems in numerical matrix analysis. This difficulty arises for two main reasons. First, deciding which eigenvalues are multiple in the presence of rounding errors is inherently problematic. Second, the determination of the sizes of the Jordan blocks associated with a given multiple eigenvalue (the Segre characteristic) is closely related to computing of the numerical rank of a matrix, which itself is a difficult task working with finite-precision arithmetic.

The Jordan canonical form is structurally unstable in the sense that it is not a continuous function of the matrix entries. The following example of a 2×2 matrix illustrates this instability.

Example 17. Let

$$A(\varepsilon) = \begin{bmatrix} 0 & 1 \\ \varepsilon & 0 \end{bmatrix},$$

where ε is a small positive number. For $\varepsilon = 0$, this matrix has the Jordan canonical form

$$J = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},$$

whereas for $\varepsilon > 0$ it has the Jordan canonical form

$$J = \begin{bmatrix} -\sqrt{\varepsilon} & 0 \\ 0 & \sqrt{\varepsilon} \end{bmatrix}.$$

Clearly, the Jordan canonical form of $A(\varepsilon)$ changes its structure discontinuously at $\varepsilon = 0$; that is, it is not continuous at this point. For small ε , the nonsingular matrix that diagonalizes $A(\varepsilon)$,

$$V = \begin{bmatrix} 1 & 1 \\ -\sqrt{\varepsilon} & \sqrt{\varepsilon} \end{bmatrix}$$

is ill-conditioned, since

$$\text{cond}_2(V) = \frac{1}{\sqrt{\varepsilon}}.$$

For example, if $\varepsilon = 10^{-8}$, then the condition number of V with respect to inversion is $\text{cond}_2(V) = 10^4$.

The eigenvalue problem for a matrix A with distinct eigenvalues λ_i , $i = 1, 2, \dots, n$ is a well-posed computational problem, since for a sufficiently small perturbation δA , the eigenvalues $\tilde{\lambda}_i$, $i = 1, 2, \dots, n$ of the perturbed matrix $A + \delta A$ satisfy the inequality

$$|\tilde{\lambda}_i - \lambda_i| \leq \kappa_{\lambda_i} \|\delta A\|_2, \quad (32)$$

where κ_{λ_i} is the condition number of λ_i (Section 3.5.2). If κ_{λ_i} is large, then the eigenvalue λ_i is ill-conditioned and in the limiting case of defective eigenvalues the condition number becomes infinite. In such cases, the eigenvalue problem is ill-posed.

In case of an ill-posed eigenvalue problem, a perturbation of magnitude ε applied to a Jordan block $J_{s_{ji}}(\lambda_i) \in \mathbb{C}^{s_{ji} \times s_{ji}}$ of order $s_{ji} > 1$ may change its eigenvalues by an amount proportional to $\varepsilon^{1/s_{ji}}$, whose derivative at $\varepsilon = 0$ is infinite (see Example 12). This extreme sensitivity is the source of the ill-posedness.

The set of matrices with defective eigenvalues forms a low-dimensional surface Π in the n^2 -dimensional parameter space of matrix entries. Consequently, matrices that lie in the vicinity of Π give rise to ill-conditioned eigenvalue problems. When the problem lies exactly on Π , it usually has bounded conditioning since the sensitivity of the defective eigenvalues is finite (see (18)). For this reason, the solution of ill-conditioned problems is often obtained by projecting them onto Π and solving the resulting ill-posed problem. Such an approach is called *regularization*.

The use of regularization in the solution of the eigenvalue problem is illustrated by the following example.

Example 18. Consider the matrix

$$A = \begin{bmatrix} -109 & -126 & 675 \\ -45 & -52 & 279 \\ -26 & -30 & 161 \end{bmatrix}.$$

Using the MATLAB[®] function `eig`, one obtains the eigenvalues

$$\begin{aligned} &1.999999999999170 + j_0.000000000000000, \\ &-0.999999999999574 + j_0.000001024409445, \\ &-0.999999999999574 - j_0.000001024409445. \end{aligned}$$

The corresponding eigenvalue condition numbers, determined by the MATLAB[®] function `condeig`, are

$$\begin{aligned} &1.6689517668 \cdot 10^2, \\ &1.3224701370 \cdot 10^8, \\ &1.3224701370 \cdot 10^8, \end{aligned}$$

which shows that the last two eigenvalues are extremely sensitive. As a consequence, the eigenvector matrix V is also ill conditioned, with condition number

$$\text{cond}(V) = 3.24 \cdot 10^8.$$

The eigenvalue problem can be regularized in the following way. The matrix A is reduced to Jordan form, using the algorithm presented in [121,122]. Within this algorithm, the last two eigenvalues are recognized as multiple and are replaced by their mean value,

$$(\lambda_2 + \lambda_3)/2 = -0.99999999999574.$$

This approach is justified by the fact that, although the individual eigenvalues may be highly sensitive, their mean value is not. In the present case, the errors in the computed second and third eigenvalues are of the order $\sqrt{\mathbf{u}\|A\|_2} \approx 1 \cdot 10^{-6}$, whereas the mean value changes by only $4.26 \cdot 10^{-13}$, which is of order of the backward error $\mathbf{u}\|A\|_2$, where \mathbf{u} denotes the unit roundoff.

In this way the ill conditioned eigenvalue problem is projected onto the set of ill-posed problems which includes the matrices with defective multiple eigenvalues. As a result of the reduction of A to Jordan form, one obtains

$$J = Z^{-1}AZ = \begin{bmatrix} 1.99999999999214 & 0 & 0 \\ 0 & -0.99999999999611 & 15.248813586295865 \\ 0 & 0 & -0.99999999999611 \end{bmatrix},$$

where

$$Z = \begin{bmatrix} -0.886460226183731 & 0.928279121632659 & 0.334637240705078 \\ -0.406294270334237 & 0.309426373877600 & -0.936984273974103 \\ -0.221615056545937 & 0.206284249251711 & -0.100391172211498 \end{bmatrix}.$$

Clearly, there is a quadratic elementary divisor corresponding to the eigenvalue -1.0 , that is, this eigenvalue belongs to a Jordan block of size 2. In the same time, the transformation matrix Z has condition number $\text{cond}(Z) = 333.79$, which is relatively modest.

In summary, due to rounding errors and the nature of standard eigenvalue algorithms, the eigenvalues are initially computed as simple but ill-conditioned. The algorithm for reduction to Jordan form correctly recognizes these eigenvalues as multiple and determines them with maximal possible accuracy. In this way, the ill-conditioned eigenvalue problem is transformed into the problem of determining multiple eigenvalues, which is then solved accurately by reducing the matrix to Jordan form.

Example 18 confirms that the computation of defective multiple eigenvalues of a matrix is an ill-posed problem. This problem can be regularized by applying an appropriate criterion to determine the dimensions of the Jordan blocks when constructing the Jordan structure of the matrix. In this way, one can determine the exact canonical form \hat{J} of a near matrix $A + \delta A$, where the norm of δA provides an upper bound on the distance of A to the regularized problem with computed structure \hat{J} .

5.4. Numerical Jordan Form

Following the presentation of Zeng [123,124] and Zeng and Li [125], the problem of determining the Jordan form of a matrix in presence of errors can be formalized as follows.

The determination of the Jordan form of a matrix with an ill-conditioned eigenvalue problem may be illustrated in a simplified setting, as shown in Figure 33 where the space \mathbb{R}^3 is used as a substitution for the n^2 dimensional parameter space of matrix entries. (Note that the ordering of the strata in the figure is purely illustrative.) The objective is to find the Jordan canonical form of the matrix A , represented as a point lying on a manifold Π .

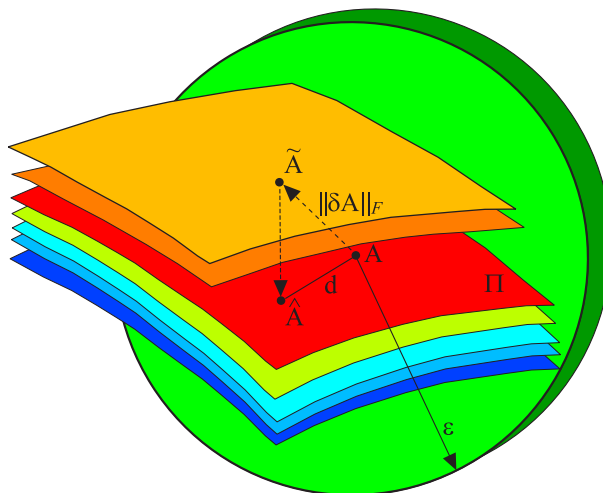


Figure 33. Determining the numerical Jordan form.

In practice, the exact matrix A is not known. Instead, one works with an approximation $\tilde{A} = A + \delta A$, which is contaminated by empirical and/or rounding errors δA satisfying $\|\delta A\| < \varepsilon$. With respect to the Frobenius norm, the point \tilde{A} lies inside a sphere of radius ε centered at A . From a theoretical point of view, the matrix \tilde{A} typically has distinct eigenvalues and it is therefore used by numerical methods to compute an approximation of the eigenstructure. However, since \tilde{A} lies outside the manifold Π , its eigenvalue problem is ill-conditioned, and the corresponding numerical results may contain large errors.

To regularize the problem, the point \tilde{A} is projected onto the manifold Π , yielding a new matrix \hat{A} whose eigenvalues are defective but with bounded sensitivity as shown by (18). Note that there exist several nonintersecting manifolds, each corresponding to a different Jordan structure. It can be shown rigorously [124], that the best regularization results are obtained when the distance

$$\text{dist}(\hat{A}, \tilde{A}) = \|\hat{A} - \tilde{A}\|$$

between \tilde{A} and Π is minimal, which corresponds to the orthogonal projection of \tilde{A} onto the closest pejorative manifold Π . This observation shows that the numerical determination of the Jordan form can be recast as a least-squares problem. As a result, the Jordan form J of \hat{A} is taken as the Jordan canonical form of A . The quantity

$$\text{dist}(\hat{A}, A) = \|\hat{A} - A\|$$

characterizes the backward error in finding J .

Thus, we arrive at the following rigorous definition of the notion of a numerical Jordan form.

Definition 17. Let $A \in \mathbb{C}^{n \times n}$ and let $\varepsilon > 0$. Suppose that an approximation

$$\tilde{A} = A + \delta A$$

of A is given, where $\|\delta A\| < \varepsilon$. Let $\Pi \subset \mathbb{C}^{n \times n}$ be a matrix bundle such that

$$\text{codim}(\Pi) = \max\{\text{codim}(\Pi') : \text{dist}(\tilde{A}, \Pi') < \varepsilon\} \quad (33)$$

and let $\hat{A} \in \Pi$ be a matrix satisfying

$$\|\hat{A} - A\|_F = \min_{B \in \Pi} \|B - A\|_F$$

with exact Jordan decomposition $\hat{A} = ZJZ^{-1}$. Then the matrix J is called the **numerical Jordan canonical form** of A within ε , and ZJZ^{-1} is called the **numerical Jordan decomposition** of A within ε .

Example 19. Consider the 6×6 matrix

$$A = \begin{bmatrix} -670 & -1302 & 4610 & 11257 & 895 & -16559 \\ -1481 & -2740 & 9987 & 24831 & 3390 & -40853 \\ 820 & 1652 & -5758 & -13884 & -556 & 18800 \\ -792 & -1525 & 5441 & 13336 & 1225 & -20142 \\ -320 & -574 & 2127 & 5345 & 905 & -9323 \\ -184 & -340 & 1240 & 3084 & 424 & -5083 \end{bmatrix}$$

whose exact Jordan canonical form is

$$J = \text{diag} \left(\begin{bmatrix} -2 & 1 & & & & \\ & -2 & 1 & & & \\ & & -2 & 1 & & \\ & & & -2 & 1 & \\ & & & & -2 & 1 \\ & & & & & -2 \end{bmatrix}, \begin{bmatrix} -1 & 1 \\ & -1 \end{bmatrix} \right).$$

This matrix is defective but nonderogatory. It belongs to a bundle Π in the 36-dimensional space whose codimension is $4 + 2 - 2 = 4$. The diagonal elements of the Schur form, computed using the function `schur` in MATLAB[®], are

$$\begin{aligned} & -2.005581021025717 + j_0 0.000000000000000, \\ & -2.000018276197160 + j_0 0.005599556627086, \\ & -2.000018276197160 - j_0 0.005599556627086, \\ & -1.994382426681778 + j_0 0.000000000000000, \\ & -0.99999999949297 + j_0 0.000018939381468, \\ & -0.99999999949297 - j_0 0.000018939381468 \end{aligned}$$

These eigenvalues are distinct and therefore correspond exactly to the eigenvalues of a nearby matrix $\tilde{A} + \delta A$, which lies in a bundle of codimension $\text{codim}(\mathcal{B}) = 0$. Using a numerical algorithm, one computes a Jordan form \hat{J} with Segre characteristics $s_{11} = 4$ and $s_{12} = 2$ associated with the eigenvalues

$$\begin{aligned} & -2.000000000025454, \\ & -2.000000000025454, \\ & -2.000000000025454, \\ & -2.000000000025454 \end{aligned}$$

and

$$\begin{aligned} & -0.99999999949297, \\ & -0.99999999949297, \end{aligned}$$

respectively. Thus, the eigenvalues of A are computed correctly to approximately eleven decimal digits. The computed Segre characteristics show that the matrix

$$\hat{A} = Z\hat{J}Z^{-1} = 10^4 \cdot \begin{array}{ccc} \left[\begin{array}{ccc} -0.06699999999759 & -0.130200000000348 & 0.461000000000600 \\ -0.148100000000135 & -0.274000000000015 & 0.998700000000243 \\ 0.082000000000064 & 0.165199999999896 & -0.575799999999809 \\ -0.079199999999947 & -0.152499999999883 & 0.544099999999599 \\ -0.031999999999793 & -0.057399999999809 & 0.212699999999166 \\ -0.018399999999934 & -0.033999999999923 & 0.123999999999690 \\ 1.125699999999035 & 0.0894999999992746 & -1.655899999978640 \\ 2.483100000001330 & 0.339000000002283 & -4.085300000008000 \\ -1.388400000000210 & -0.055600000002015 & 1.880000000005861 \\ 1.333599999999064 & 0.122500000000053 & -2.014199999998984 \\ 0.534499999997335 & 0.090499999997931 & -0.932299999990937 \\ 0.308399999999091 & 0.042399999999476 & -0.508299999997412 \end{array} \right] \end{array} \begin{array}{l} \dots \rightarrow \\ \left. \dots \leftarrow \right. \end{array}$$

belongs to the same bundle Π as A , with codimension equal to 4. The exact Jordan canonical form of \hat{A} is \hat{J} .

The relative distance between \hat{A} and A , which characterizes the backward error in computing \hat{J} , is

$$\text{dist}(\hat{A}, A) = \|\hat{A} - A\|_F / \|A\|_F = 4.16 \cdot 10^{-12}.$$

Note that in some cases the matrix \hat{A} may lie in a different bundle from A , if the Segre characteristics of A are not identified correctly.

It should be noted that, in contrast to the “theoretical case”, the numerical Jordan structure remains unchanged within a certain set of parameter values defined by the inequality

$$\|E\|_F \leq \varepsilon. \quad (34)$$

The parameter set determined by (34), is represented as a ball of radius ε , centered at the singular point corresponding to the theoretical case. Thus, unlike the exact Jordan form, which is defined for a single combination of parameters at a singular point, the numerical Jordan form remains the same for all parameter values contained in the ball of radius ε .

Efficient numerical algorithm for computing the numerical Jordan form of a matrix is presented in [121,122,126,127], see also [128].

6. Matrices Depending on Parameters

6.1. Matrix Deformations

In physical problems, the entries of a matrix A often depend on certain parameters. Suppose these parameters belong to a parameter space \mathcal{P}^k with $\dim(\mathcal{P}^k) = k$, where k is the number of independent parameters. Let $A_0 \in \mathbb{C}^{m \times n}$ be a fixed matrix. A family of matrices

$$A(\delta), \delta \in \Delta \subseteq \mathcal{P}^k,$$

where Δ is a neighborhood of the origin in \mathcal{P}^k , is called a *deformation*, of A_0 if the mapping

$$A : \Delta \rightarrow \mathbb{C}^{m \times n}$$

is such that each entry of $A(\delta)$ is a convergent power series in the δ_i , and $A(0) = A_0$. A deformation is also called *matrix family* and the subset $\Delta \subseteq \mathcal{P}^k$ is referred to as the *base* of the family.

A function that is locally given by a convergent power series is called *analytic*, and the complex case, it is called *holomorphic*. A holomorphic function is infinitely differentiable at every point in its domain and is therefore continuous. A mapping $A : \Delta \rightarrow \mathbb{C}^{m \times n}$ whose entries are holomorphic functions is called *holomorphic mapping*. Since the entries depend smoothly on the parameters, such mappings are convenient for computational implementation.

When a matrix depends on parameters, we say that it is given a *family of matrices*. In practice, we are usually interested in the family locally, i.e., for small changes of the parameters near fixed values. In such cases, we speak of *deformations* of the matrix corresponding to these small parameter changes.

In Figure 34 we symbolically represent a k -parameter generic matrix family that intersects the variety of singular cases transversally (i.e., the intersection occurs at “nonzero angle”). Matrix families that are transverse to all varieties are called *generic families*.

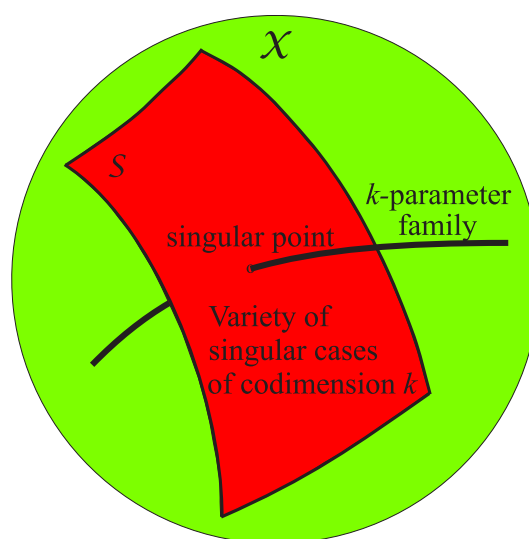


Figure 34. A generic k -parameter family and the variety of singular cases.

A family with k parameters can be viewed as a k -dimensional manifold in the matrix space. For instance, a one-parameter family is represented by a curve in a 3-dimensional matrix space with codimension equal to 2, while the variety of singular cases has codimension equal to 1.

The variety of singular cases depends on the specific problem. In the context of solving linear systems of equations or inverting matrices, this variety is the set of all singular matrices. For the eigenvalue problem, the variety of singular cases consists of all matrices that are defective or/and derogatory. In this setting, the one-dimensional family typically contains matrices with simple eigenvalues except at the singular points, where the curve representing the one-parameter family intersects the manifold of singular cases.

By the principle of transversality, in the general case, the variety of singular cases has dimension $n \cdot m - k$ and a codimension equal to k . Hence, the codimension of the variety of singular cases is equal to the number of parameters that determine the matrix family.

We now illustrate this concept with an example related to the eigenvalue problem.

Example 20. Consider a two-parameter deformation of the 2×2 Jordan block

$$A_0 = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}$$

given by

$$A(x, y) = A_0 + \Delta(x, y), \quad \Delta(x, y) = \begin{bmatrix} x + 10x^2 + y + 6y^2 & 0 \\ 2x^2 + 3y^2 & -5x + 7y \end{bmatrix}.$$

In this example, the orbit of A_0 has dimension 2, and the tangent space of the perturbation $\Delta(x, y)$ also has dimension 2 for each pair x, y . Consequently,

$$\dim(\mathcal{O}(A_0)) + \dim(\mathcal{O}(\Delta(x, y))) = n^2,$$

i.e., the perturbation $\Delta(x, y)$ is in general position with respect to the orbit of A_0 for all x, y .

Figure 35 shows the matrix family $A(x, y)$ for $x \in [-0.2, 0.2]$, $y \in [-0.2, 0.2]$. The entries

$$a_{11} = -1, a_{21} = 0, a_{22} = -1$$

correspond to a singular point at which A has double eigenvalue and canonical form which is a 2×2 Jordan block A_0 . By introducing infinitesimal changes in x and y , the double eigenvalue splits into two simple eigenvalues, and the Jordan form becomes diagonal.

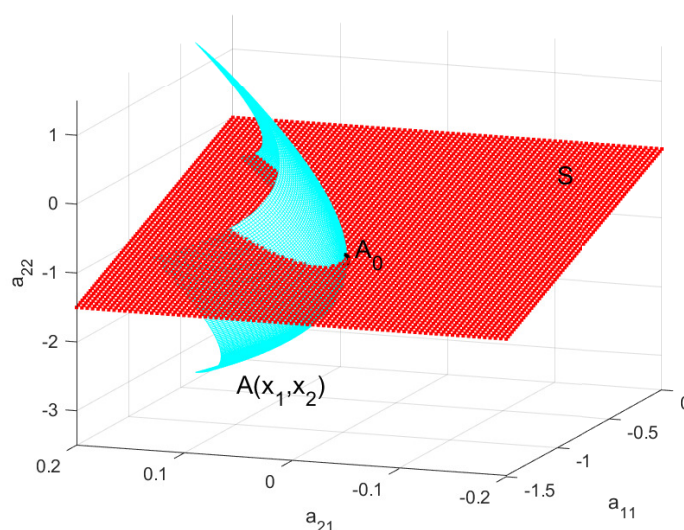


Figure 35. Two-parameter matrix deformation and the manifold of singular cases.

The singular points corresponding to matrices with multiple eigenvalues in the form

$$A = \begin{bmatrix} \mu & v \\ 0 & \mu \end{bmatrix}, v \neq 0,$$

lie on the variety of the singular cases S , which in this example represents a plane in the three-dimensional parameter space.

6.2. Versal Deformations

We shall adopt the following terminology. By a map $B(\eta) \rightarrow A(\delta)$ of one family into another family, we mean a correspondence in which to each value of the parameter η of the family $B(\eta)$ there corresponds a definite value $\delta = \varphi(\eta)$ of the parameter of the family $A(\delta)$.

A **versal family** of matrices is one into which we can map every other family of matrices by means of a suitable mapping.

A **universal family** is a versal family $A(\delta)$ with the additional property that in mapping any family $B(\eta)$ into it, the “change of parameters” $\delta = \varphi(\eta)$ is uniquely determined by $B(\eta)$.

A **miniversal family** is a versal family depending on the minimum possible number of parameters. Clearly, such families are of particular interest from a computational point of view.

The term “versal” is formed from the word “universal” by dropping the prefix “uni”, which signifies the uniqueness of the map φ . As Arnold remarks [129], “versal” is the intersection of the concepts of “universal” and “transversal”.

To formalize these notations rigorously, we introduce the following definitions.

Two deformations $A(\delta)$ and $A'(\delta)$ of the $m \times n$ matrix A_0 are called **equivalent** if there exist deformations $P(\delta)$ and $Q(\delta)$ of the identity matrices I_m and I_n , respectively, both defined on the same base Δ , such that

$$A'(\delta) = P(\delta)A(\delta)Q^{-1}(\delta).$$

In other words, the deformation $A'(\delta)$ is obtained from $A(\delta)$ via an equivalent matrix transformation.

Let $A(\delta)$ and $A'(\eta)$ be deformations of A_0 , where $\delta \in \Delta \subseteq \mathcal{P}^k$ and $\eta \in H \subseteq \mathcal{P}^\ell$, with \mathcal{P}^k and \mathcal{P}^ℓ being parameter spaces of dimensions k and ℓ , respectively.

If there exist deformations $P(\eta)$ of the identity matrix I_m and $Q(\eta)$ of the identity matrix I_n , defined for $\eta \in \tilde{H} \subseteq H$, together with a holomorphic mapping $\varphi : \tilde{H} \rightarrow \Delta$ satisfying $\varphi(0) = 0$, such that

$$A'(\eta) = P(\eta)A(\varphi(\eta))Q^{-1}(\eta), \quad \eta \in \tilde{H},$$

then we say that $A'(\eta)$ is **induced** from the deformation $A(\delta)$ via $P(\eta)$, $Q(\eta)$ and $\varphi(\eta)$ (see Figure 36).

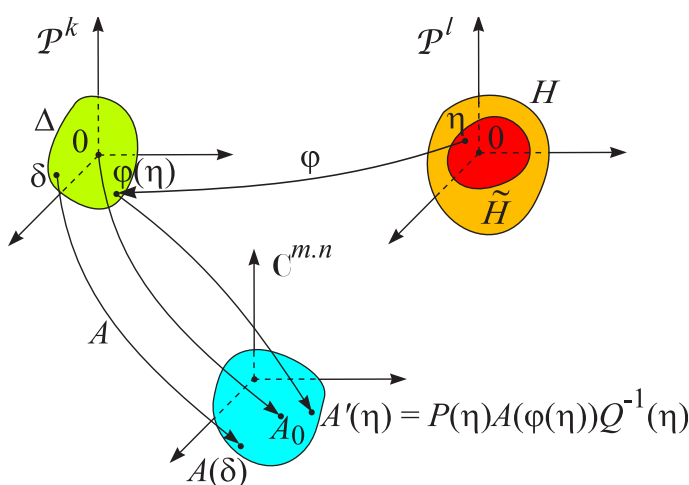


Figure 36. Versal deformation.

A deformation $A(\delta)$ of a matrix A_0 is said to be **versal**, if every other deformation $A'(\eta)$ of A_0 is equivalent to a deformation induced from $A(\delta)$ via a suitable parameter change. That is, there exist deformations $P(\eta)$ of I_m and $Q(\eta)$ of I_n , and a holomorphic mapping $\varphi(\eta)$ with

$$P(0) = I_m, \quad Q(0) = I_n, \quad \varphi(0) = 0,$$

such that

$$A'(\eta) = P(\eta)A(\varphi(\eta))Q^{-1}(\eta), \quad P(0) = I_m, \quad Q(0) = I_n, \quad \varphi(0) = 0. \tag{35}$$

A versal deformation of A_0 is called **universal**, if the inducing map φ is uniquely determined by A' .

A versal deformation of A_0 is called **miniversal** if its parameter space has minimal dimension among all versal deformations of A_0 .

We now proceed to a characterization of versal deformations.

The following theorem gives a condition under which a matrix deformation is versal.

Theorem 18. [100] *A deformation $A(\delta)$ of A_0 is a versal deformation if and only if $A(\delta)$ is transversal to the orbit of A_0 at $\delta = 0$.*

We now show that a versal deformation $A(\delta)$ is indeed transversal. Let $A'(\eta)$ be an arbitrary deformation of A_0 . By the versality of $A(\delta)$, there exist deformations $P(\eta), Q(\eta)$, and a mapping $\varphi(\eta)$ such that

$$A'(\eta) = P(\eta)A(\varphi(\eta))Q^{-1}(\eta), \quad \eta \in \tilde{H}.$$

Differentiating and taking into account that

$$\frac{\partial Q^{-1}}{\partial \eta} = -Q^{-1} \frac{\partial Q}{\partial \eta} Q^{-1}, \quad Q(0) = Q^{-1}(0) = I_n$$

we obtain

$$A'_* = A_*\varphi_* + P_*A_0 - A_0Q_*,$$

where the subscript $*$ denotes differentiation with respect to η at $\eta = 0$.

Consequently, for every tangent vector $v \in \mathcal{P}^\ell$ at the base of A' , we have that

$$A'_*v = A_*\varphi_*v + (P_*A_0 - A_0Q_*)v.$$

By Lemma 1, the vector $(P_*A_0 - A_0Q_*)v$ belongs to the tangent space of the orbit $\mathcal{O}(A_0)$. Therefore, any vector in the tangent space $T_{A_0}A(\delta)$ at A_0 can be expressed as the sum of a vector in the image of A_* and a vector tangent to the orbit of A_0 .

Hence, $A(\delta)$ is transversal to the orbit of A_0 , as illustrated in Figure 37.

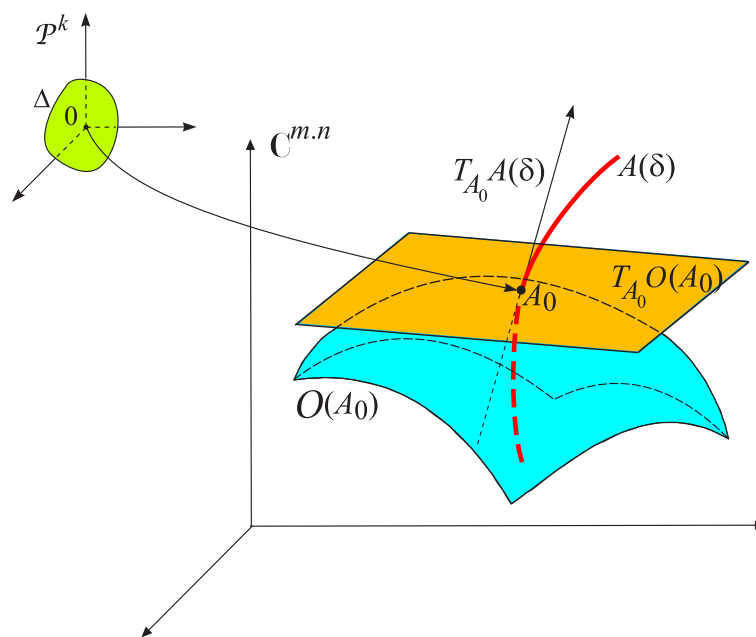


Figure 37. The versal deformation $A(\delta)$ is transversal to $\mathcal{O}(A_0)$.

A proof that a deformation which is transversal to the orbit is versal, can be found in ([113], Sect. 2.9).

According to Theorem 18, we have

$$T_{A_0}A(\delta) + T_{A_0}\mathcal{O}(A_0) = C^{m \times n}. \quad (36)$$

This implies that the dimension of the parameter space of a versal deformation of A_0 is equal to

$$\dim(\Delta) = m \cdot n - \dim(\mathcal{O}(A_0)),$$

i.e., it equals the codimension of the orbit $\mathcal{O}(A_0)$.

Equation (36) represents a particular case of a general situation, which can be described as follows. Let $N \subset M$ be a smooth submanifold of a manifold M . Consider a mapping

$$A : \Delta \rightarrow M$$

of another manifold Δ into M , and let $\delta \in \Delta$ be such that $A(\delta) \in N$.

The mapping $A : \Delta \rightarrow M$ is said to be transversal to $N \subset M$ at δ if the tangent space to M at $A(\delta)$ is the sum of the image of the differential of A and the tangent space to N , i.e.,

$$T_{A(\delta)}\mathcal{M} = DA(T_\delta\Delta) + T_{A(\delta)}\mathcal{N}. \tag{37}$$

Equation (37) gives a condition for a transversal intersection of the manifolds Δ and \mathcal{N} (Figure 38).

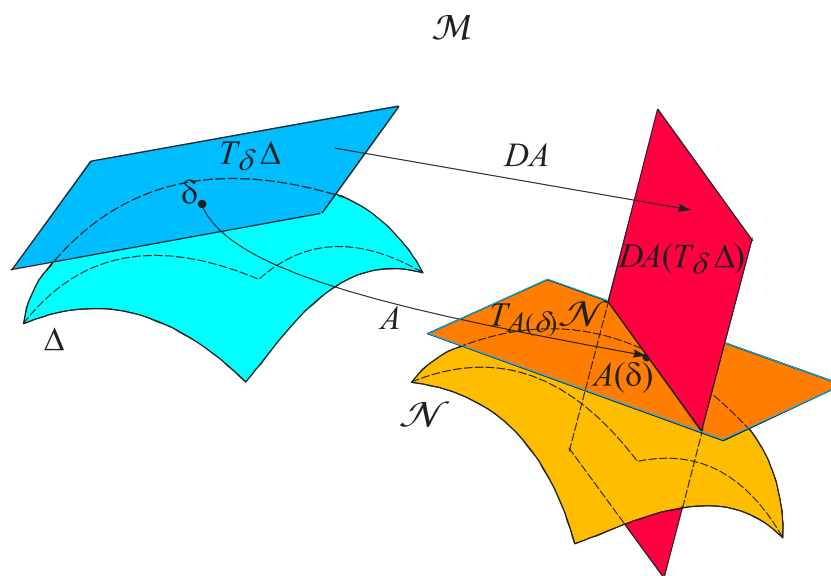


Figure 38. The mapping $A : \Delta \rightarrow M$ is transverse to the manifold \mathcal{N} .

Example 21. (a) Let

$$A_0 = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}, \lambda_i \neq \lambda_j \text{ for } i \neq j.$$

Then the codimension of the orbit $\mathcal{O}(A_0)$ is equal to n , and an n -parameter versal deformation of A_0 is

$$A(\delta) = \begin{bmatrix} \lambda_1 + \delta_1 & & & \\ & \lambda_2 + \delta_2 & & \\ & & \ddots & \\ & & & \lambda_n + \delta_n \end{bmatrix},$$

where $\delta_i, i = 1, 2, \dots, n$ are arbitrary parameters. This deformation is both universal and miniversal.

(b) Let $A_0 = 0$. Then $\text{codim}(\mathcal{O}(A_0)) = n^2$, and an n^2 -parameter versal deformation of A_0 is

$$A(\delta) = \begin{bmatrix} \delta_{11} & \delta_{12} & \dots & \delta_{1n} \\ \delta_{21} & \delta_{22} & \dots & \delta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{n1} & \delta_{n2} & \dots & \delta_{nn} \end{bmatrix},$$

that is, the family of all $n \times n$ matrices. This deformation is also miniversal.

Equation (35) may be interpreted as a local approximation, in a neighbourhood of the origin, of an arbitrary matrix family $A(\delta)$ by a versal family $A'(\eta)$. A versal deformation of a matrix A_0 thus plays the role of a “normal form” into which not only the single matrix A_0 , but also any family of matrices sufficiently close to A_0 , can be transformed.

Naturally, this normal form must itself depend on parameters. Its principal advantage is that both the entries of the normal form and the similarity transformation leading to it can be chosen to depend smoothly on the entries of the original matrix, as they vary in a neighbourhood of A_0 .

Moreover, provided that the second order terms neglected in the linear approximation (26) are sufficiently small, versal deformations preserve the bundle containing A_0 , that is, the matrices of the family $A(\delta)$ remain in the same bundle as A_0 .

Further details on the versal deformations of matrices can be found in [117,119,130]. Such deformations are used to construct normal forms of square matrices that depend smoothly on parameters [100].

6.3. Bifurcation Diagrams

Consider the structure and properties of the set in the parameter space, corresponding to the variety of singular cases in the matrix space associated with the eigenvalue problem. In this case, the parameter space can be partitioned into subsets, correspondingly to the partition of the matrix space in bundles. The exceptional values of the parameters to which correspond matrices with multiple eigenvalues (the singular cases), constitute a subset in the parameter space. This subset is called **bifurcation diagram**. The bifurcation diagram of a generic family of matrices represents a finite union of varieties – to each bundle of orbits corresponds its own variety in the parameter space \mathcal{P} . The codimension k of a variety in the parameter space of a generic family is equal to the codimension of the corresponding bundle in the space of all matrices, i.e., to the number of parameters determining the matrix family. Therefore, the bifurcation diagram of a family of matrices is a partition of the parameter space \mathcal{P} according to Jordan types of matrices. The partition consists of grouping together the matrices with the same dimensions of the Jordan blocks differing only in the eigenvalues. Thus, the bifurcation diagrams allow studying the partition of space of matrices into matrices with Jordan forms of distinct types. This makes them a useful tool to analyze the qualitative metamorphosis (or “catastrophes”) of a matrix family.

The partition of the matrix space into matrices with the same dimensions of the Jordan blocks groups the matrices into bundles of equal codimensions and represents a finite stratification of the space of matrices. In the space of families of matrices of order n , the families transversal to the stratification into Jordan types constitute an everywhere dense set. Clearly, such families consist of matrices with simple eigenvalues.

The term “bifurcation” was introduced by Poincaré in 1885 in a paper that marks the beginning of bifurcation theory [131].

An accessible introduction to the bifurcation theory of dynamical systems is given in [132], while a more comprehensive and in-depth treatment can be found in [22,133,134]. A bifurcation analysis of eigenvalues and generalized eigenvalues is presented in [135], Ch. 2.

Below, we consider bifurcation diagrams of two-, and three-parameter matrix families, associated with strata containing Jordan blocks of different sizes.

Example 22. Consider a family of third order matrices in companion form

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_3 & -a_2 & -a_1 \end{bmatrix}$$

with characteristic equation

$$z^3 + a_1z^2 + a_2z + a_3 = 0.$$

Using the substitution

$$z = t - a_1/3,$$

we obtain the depressed cubic equation

$$t^3 + pt + q = 0, \quad (38)$$

$$p = a_2 - a_1^2/3, \quad q = 2a_1^3/27 - a_1a_2/3 + a_3.$$

Representation (38) shows that the parametric space associated with the matrix bundle under consideration is two-dimensional, coinciding with the codimension of the bundle.

In the generic case, the discriminant

$$\Delta = -(4p^3 + 27q^2)$$

is nonzero, and the characteristic equation has three disjoint roots α, β, γ . In this case, the Jordan form of A is diagonal,

$$J = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{bmatrix}.$$

This stratum will be denoted by $\alpha\beta\gamma$.

If the discriminant $\Delta = 0$ and $p \neq 0$, then the cubic equation has a double root

$$\alpha = \alpha_1 = \alpha_2 = -\frac{3q}{2p} - \frac{a_1}{3} = \frac{9a_3 - a_1a_2}{2(a_1^2 - 3a_2)}$$

and a simple root

$$\beta = -\frac{3q}{p} - \frac{a_1}{3} = \frac{4a_1a_2 - 9a_3 - a_1^3}{a_1^2 - 3a_2}.$$

The Jordan form is then

$$J = \left[\begin{array}{cc|c} \alpha & 1 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \beta \end{array} \right]$$

with a 2×2 Jordan block corresponding to the double eigenvalue α and a 1×1 block corresponding to the simple eigenvalue β . This stratum is denoted by $\alpha^2\beta$.

If $\Delta = 0$ and $p = 0$, then $p = q = 0$ and

$$\alpha = \alpha_1 = \alpha_2 = \alpha_3 = -\frac{a_1}{3}$$

is a triple eigenvalue of A . Since the companion matrices are non-derogatory, the triple eigenvalue participates in one Jordan block:

$$J = \begin{bmatrix} \alpha & 1 & 0 \\ 0 & \alpha & 1 \\ 0 & 0 & \alpha \end{bmatrix}.$$

This stratum is denoted by α^3 .

In Figure 39 we show the discriminant Δ of the characteristic equation of a 3×3 companion matrix for $a_1 = 1$ and various values of the coefficients a_2 and a_3 . The bifurcation diagram of the matrix bundle is obtained as the intersection of the discriminant surface with the plane $\Delta = 0$. The bifurcation diagram forms a semi-cubic parabola—a curve with a singular point in the shape of a cusp, which corresponds to the triple eigenvalue α (Figure 40).

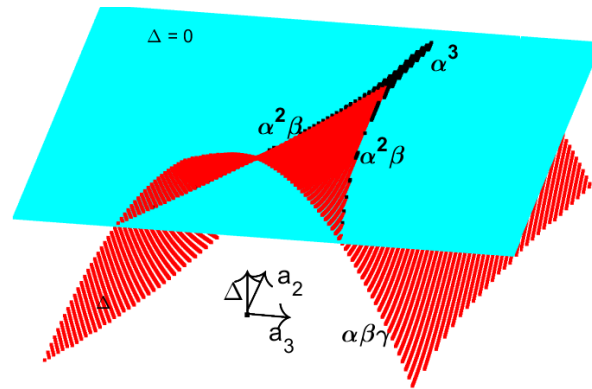


Figure 39. Bifurcation diagram for 3rd order companion matrices.

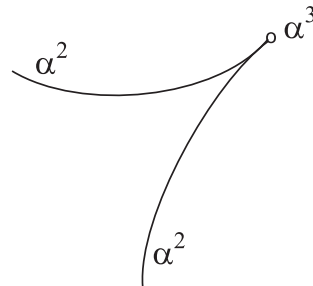


Figure 40. Bifurcation diagram for α^3 in the parameter plane.

Example 23. Consider a family of 2×2 matrices

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

In the regular (generic) case, such matrices have the diagonal Jordan form

$$J = \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix},$$

where α and β are simple eigenvalues. For simplicity, a stratum of this type will be denoted by $\alpha\beta$.

We are interested in the degenerate case when the Jordan form

$$J = \begin{bmatrix} \alpha & 0 \\ 0 & \alpha \end{bmatrix}$$

consists of two 1×1 blocks and the double eigenvalue α is semisimple. This stratum will be denoted by $\alpha\alpha$. The matrices $PJP^{-1} = \alpha I_2$ are non-defective but derogatory. In the given case, the eigenvalue $\alpha = -a_1/2$ is determined as the double root of the quadratic equation $z^2 + a_1z + a_2 = 0$ for values of the parameters a_1 and a_2 zeroing the discriminant,

$$\Delta = a_1^2 - 4a_2 = 0. \tag{39}$$

Equation (39) is parameterized as

$$\sqrt{x_1^2 + x_2^2} = \pm r, \tag{40}$$

where

$$x_1 = r \cos(\Theta), x_2 = r \sin(\Theta), 0 \leq \Theta \leq 2\pi, r = 2\sqrt{a_2}.$$

The expression (40) describes a variety of singular cases in \mathbb{C}^3 in the form of a cone with vertex at $r = 0$ (Figure 41). The point $r = 0$, corresponding to the double semisimple eigenvalue α , is a singular point of the variety, while the conical surfaces for $r \neq 0$ corresponds to bundles of type α^2 with codimension 1 (single Jordan block of order 2). The points $\alpha\beta$ outside the bifurcation diagram represent matrices with two distinct simple eigenvalues α and β .

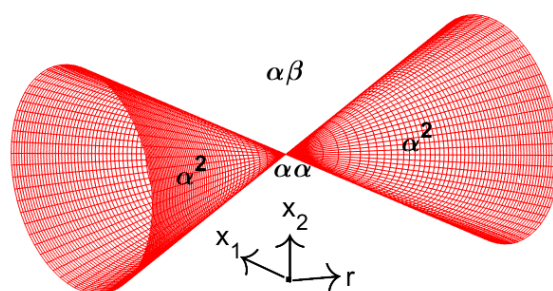


Figure 41. Bifurcation diagram for matrices with double eigenvalue.

Example 24. Consider a three-parameter matrix family consisting of 4×4 companion matrices of the form

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -e & -d & -c & 0 \end{bmatrix}.$$

The singular cases of this family consist of matrices that are defective and non-derogatory. Note that a fourth order companion matrix with a quadruple eigenvalue has only one Jordan block in its Jordan canonical form.

The characteristic equation of the matrix A has the form of a depressed quartic polynomial equation

$$z^4 + cz^2 + dz + e = 0.$$

This equation has repeated roots if and only if

$$\Delta(c, d, e) = 0, \quad (41)$$

where

$$\Delta(c, d, e) = 16c^4e - 4c^3d^2 - 128c^2e^2 + 144cd^2e - 27d^4 + 256e^3$$

is the discriminant of the quartic polynomial $z^4 + cz^2 + dz + e$.

Equation (41) can be written as a biquadratic equation

$$a_1d^4 + a_2d^2 + a_3 = 0$$

with respect to the parameter d , where

$$\begin{aligned} a_1 &= -27, \\ a_2 &= 144ce - 4c^3, \\ a_3 &= 16c^4e - 128c^2e^2 + 256e^3. \end{aligned}$$

This equation has four roots given by

$$d = \pm \sqrt{(-a_2 \pm \sqrt{a_2^2 - 4a_1a_3}) / (2a_1)}.$$

Thus, the bifurcation diagram corresponding to repeated roots depends on the two free parameters c and e . The surface representing the points (c, d, e) shown in Figure 42, is called a **swallowtail**.

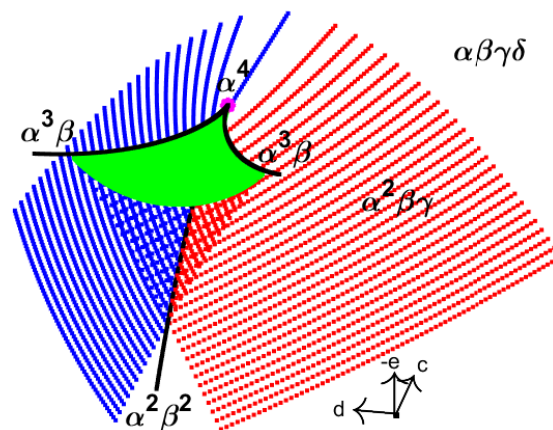


Figure 42. Swallowtail diagram for matrices of type α^4 .

According to the data presented in Table 3, the point α^4 (the swallowtail point) represents the 4×4 Jordan blocks $J_4(\alpha)$ of defective matrices, corresponding to the most degenerate companion matrices A .

Table 3. Bundles of Jordan forms represented by the swallowtail diagram

Notation	Segre characteristics	$\text{codim}(\mathcal{B})$	$\text{dim}(\mathcal{B})$	Representation in \mathbb{C}^3
α^4	{4}	3	0	point
$\alpha^3\beta$	{3}, {1}	2	1	curve
$\alpha^2\beta^2$	{2}, {2}	2	1	curve
$\alpha^2\beta\gamma$	{2}, {1}, {1}	1	2	surface
$\alpha\beta\gamma\delta$	{1}, {1}, {1}, {1}	0	3	complement of the swallowtail

The curve $\alpha^3\beta$, consisting of two cuspidal edges emanating from the swallowtail point, corresponds to the Jordan forms $J_3(\alpha) \oplus J_1(\beta)$. The curve $\alpha^2\beta^2$, given by the intersection of the swallowtail wings, represents the Jordan forms $J_2(\alpha) \oplus J_2(\beta)$.

The surface $\alpha^2\beta\gamma$, known as the swallowtail surface, represents the Jordan forms $J_2(\alpha) \oplus J_1(\beta) \oplus J_1(\gamma)$. Finally, the points $\alpha\beta\gamma\delta$, corresponding to the region outside the swallowtail, represent the diagonal Jordan forms

$$J_1(\alpha) \oplus J_1(\beta) \oplus J_1(\gamma) \oplus J_1(\delta),$$

that is, matrices with four simple eigenvalues.

The latest case is the most generic one and therefore occurs with the highest probability.

7. Conclusions

The brief survey of the applications of differential topology in matrix analysis shows that the introduction of the concepts of smooth manifolds and smooth mappings enables the analysis of various global properties of parameter-dependent matrices, including the genericity of the problems considered, the identification of singular cases, the estimation of the distance to ill-posed problems, the conditioning of problems in both deterministic and probabilistic terms, and the determination of numerical solutions as a problem for finding an optimal bundle containing an approximate solution.

Methods of differential topology are particularly well suited for matrices whose entries are only approximately known and for computations subject to rounding errors. All of this provides reason to expect that differential topology methods will find wider applications in matrix analysis.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated during the current study are available from the author upon reasonable request.

Conflicts of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

\mathbb{F} ,	number field;
\mathbb{N} ,	set of natural numbers $1, 2, \dots, n, \dots$;
\mathbb{Z} ,	set of integers $0, \pm 1, \pm 2, \dots, \pm n, \dots$;
\mathbb{Q} ,	set of rational numbers p/q , where p and q are integers;
\mathbb{R} ,	set of real numbers;
\mathbb{R}^+ ,	set of positive real numbers;
\mathbb{C} ,	set of complex numbers;
$j_0 = \sqrt{-1}$,	imaginary unit;
$A \cup B$,	union of sets A and B ;
$A \cap B$,	intersection of sets A and B ;
$A \setminus B$,	difference of sets A and B ;
$A \times B$,	Cartesian product of sets A and B ;
C^∞ ,	class of smooth functions;
C^ω ,	class of analytic functions;
$\langle x, y \rangle$,	scalar product of the vectors x and y ;
$\mathbb{C}^{n \times m}$,	space of $n \times m$ complex matrices;
$\mathbb{R}^{n \times m}$,	space of $n \times m$ real matrices;
$ A $,	matrix of absolute values of the entries of A ;

A^T ,	transposed of A ;
A^H ,	Hermitian conjugate of A ;
A^{-1} ,	inverse of A ;
$0_{m \times n}$,	$m \times n$ zero matrix;
I_n ,	$n \times n$ identity matrix;
$\det(A)$,	determinant of A ;
$\text{rank}(A)$,	rank of A ;
$\text{tr}(A)$,	trace of a square matrix A ;
$\sigma_i(A)$,	i th singular value of A ;
δA ,	perturbation of A ;
$\ x\ _2$,	Euclidean norm of the vector x ;
$\ A\ _2$,	spectral norm of A ;
$\ A\ _F$,	Frobenius norm of A ;
\mathcal{M} ,	manifold;
$\dim(\mathcal{M})$,	dimension of the manifold \mathcal{M} ;
$\text{codim}(\mathcal{M})$,	codimension of the manifold \mathcal{M} ;
$\text{vol}(\mathcal{M})$,	volume of the manifold \mathcal{M} ;
$f: \mathcal{M} \rightarrow \mathcal{P}$,	mapping of the manifold \mathcal{M} into the manifold \mathcal{P} ;
$i_{\mathcal{M}}: \mathcal{M} \rightarrow \mathcal{M}$,	identity map of \mathcal{M} into \mathcal{M} ,
$D_x f$,	differential of the map f at the point x ;
$g \circ f$,	composition of the maps g and f ;
$\mathcal{O}(A)$,	orbit of the matrix A ;
$\mathcal{B}(A)$,	bundle of the matrix A ;
$T_x \mathcal{M}$,	tangent space of \mathcal{M} at the point x ;
$T\mathcal{M}$,	tangent bundle of \mathcal{M} ;
$N_x \mathcal{M}$,	normal space of \mathcal{M} at the point x ;
$N\mathcal{M}$,	normal bundle of \mathcal{M} ;
\mathbb{S}^n ,	n -dimensional sphere in \mathbb{R}^{n+1} ;
\mathbb{T}^n ,	n -dimensional torus in \mathbb{R}^{n+1} ;
$B_r(p)$,	open ball centered at p with a radius r
$\bar{B}_r[p]$,	closed ball centered at p with a radius r ;
$\mathcal{X} = \mathcal{R}(X)$,	subspace spanned by the columns of X ;
$\mathcal{P}\{\mathbf{y} > y\}$,	probability of the event $\{\mathbf{y} > y\}$;
$E\{\xi\}$,	expected value (mean) of the random variable ξ ;
$v \perp S$,	vector v orthogonal to a subspace S ;
$:=$,	equal by definition.

References

1. H. Weyl. *Philosophy of Mathematics and Natural Science*. Princeton University Press, Princeton, NJ, 2009. ISBN: 9780691141206
2. M. Morse. *The Calculus of Variations in the Large*. AMS, New York, 1934.
3. The MathWorks, Inc., Natick, MA. *MATLAB Version 9.9.0.1538559 (R2020b)*, 2020. <http://www.mathworks.com>
4. V. Guillemin and A. Polack. *Differential Topology*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1974. ISBN: 0-13-212605-2
5. J.M. Lee. *Introduction to Topological Manifolds*, volume 202 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 2011. doi: 10.1007/978-1-4419-7940-7
6. J.M. Lee. *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 2013. doi: 10.1007/978-1-4419-9982-5
7. V.I. Arnold. *Geometrical Methods in the Theory of Ordinary Differential Equations*. Springer-Verlag, New York, second edition, 1988. ISBN: 978-1-4612-6994-6
8. V.I. Arnol'd. *Ordinary Differential Equations*. Springer-Verlag, Berlin, 1992. ISBN: 3-540-54813-0
9. L.W. Tu. *An Introduction to Manifolds*. Springer-Verlag, New York, 2011. doi: 10.1007/978-1-4419-7400-6
10. K. Burns and M. Guidea. *Differential Geometry and Topology. With a View to Dynamical Systems*. Chapman & Hall/CRC, Boca Raton, FL, 2005. ISBN: 978-1584882534

11. J.W. Milnor. *Topology from the Differential Viewpoint*. The University Press of Virginia, Charlottesville, VA, 1965.
12. A.H. Wallace. *Differential Topology. First Steps*. W.A. Benjamin, Inc., New York, 1968.
13. M.W. Hirsh. *Differential Topology*. Springer-Verlag, New York, 1976. doi: 10.1007/978-1-4684-9449-5
14. V.I. Arnold. *Mathematical Methods of Classical Mechanics*. Springer-Verlag, New York, second edition, 1989. ISBN: 0-387-96890-3
15. Y. I. Manin. The notion of dimension in geometry and algebra. *Bull. Amer. Math. Soc. (N.S.)*, 43:139–161, 2006. doi: 10.1090/S0273-0979-06-01081-0
16. E. Scholz. The concept of manifold, 1850–1950. In I.M. James, editor, *History of Topology*, pages 25–64. Elsevier, Amsterdam, 1999. ISBN: 0-444-82375-1.
17. B. Chow and C. Chow. *Lectures on Differential Geometry*. American Mathematical Society, Providence, RI, 2024. ISBN: 978-1-4704-7804-9
18. Yu. Borisovich and N. Bliznyakov and Ya. Izrailevich and T. Fomenko. *Introduction to Topology*. Mir Publishers, Moscow, 1985.
19. V.I. Arnold, V.V. Goryunov, O.V. Lyashko, and V.A. Vasil'ev. *Dynamical Systems VIII. Singularity Theory II, Applications*, volume 39 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, 1993. doi: 10.1007/978-3-662-06798-7
20. V.I. Arnold, V.V. Goryunov, O.V. Lyashko, and V.A. Vasil'ev. *Singularity Theory I*, volume 6 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, 1998. doi: 10.1007/978-3-642-58009-3
21. M. Golubitsky and V. Guillemin. *Stable Mappings and Their Singularities*. Springer-Verlag, New York, 1973. ISBN: 0-387-90073-3-X
22. J. Guckenheimer. Bifurcations of dynamical systems. In C. Marchioro, editor, *Dynamical Systems*, volume 78, pages 7–123. Springer, Berlin, 2010. doi: 10.1007/978-3-642-13929-1_1
23. R. Thom. Structural stability, catastrophe theory, and applied mathematics. *SIAM Rev.*, 19:189–201, 1977. doi: 10.1137/1019036
24. E.C. Zeeman. Catastrophe theory. In W. Güttinger and H. Eikemeier, editors, *Structural Stability in Physics*, volume 4 of *Springer Series in Synergetics*, pages 12–22. Springer, Berlin, 1979. doi: 10.1007/978-3-642-67363-4_3
25. V.I. Arnold. *Catastrophe Theory*. Springer-Verlag, Berlin, 1992. ISBN: 978-3-540-54811-9 doi:10.1007/978-3-642-58124-3
26. V.I. Arnol'd, editor. *Dynamical Systems V. Bifurcation Theory and Catastrophe Theory*, volume 5 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, 1994. doi: 10.1007/978-3-642-57884-7
27. M. Golubitsky. An introduction to catastrophe theory and its applications. *SIAM Rev.*, 20:352–387, 1978. doi: 10.1137/1020043
28. J. Guckenheimer. The catastrophe controversy. *Math. Intelligencer*, 1:15–20, 1978. doi: 10.1007/BF03023037
29. F. Chaitin-Chatelin, V. Frayssé, and T. Braconnier. Computations in the neighbourhood of algebraic singularities. *Numer. Funct. Anal. Optim.*, 16:287–302, 1995. doi: 10.1080/01630569508816619
30. C.T.C. Wall. *Differential Topology*. Cambridge University Press, Cambridge, UK, 2016. ISBN: 978-1-107-15352-3
31. R. Gilmore. Lie groups: General theory. In J.P. Francoise, G.L. Naber, and T. S. Tsun, editors, *Encyclopedia of Mathematical Physics*, pages 286—304. Elsevier, Amsterdam, 2006. doi: 10.1016/B0-12-512666-2/00425-9
32. J.B. Carrell. *Groups, Matrices, and Vector Spaces*. Springer, New York, 2017. ISBN: 978-0-387-79427-3
33. P.J. Olver. *Applications of Lie Groups to Differential Equations*. Springer-Verlag, New York, 1993. ISBN: 978-0-387-95000-6
34. K. Kendig. *Elementary Algebraic Geometry*. Springer-Verlag, New York, 1977. ISBN: 978-1-4615-6901-5
35. L. Kronecker. Über Schaaren von quadratischen und bilinearen formen. *Monatsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin vom Jahre 1874*, pages S. 59–76, 149–156, 206–232, 1874. = Werke, 1 (1895), Teubner, Leipzig, pp. 349–413. doi: 10.3931/e-rara-17875
36. J. Hadamard. *Lectures on Cauchy's Problem in Linear Partial Differential Equations*. Yale University Press, New Haven, 1923.
37. J.R. Rice. A theory of condition. *SIAM J. Numer. Anal.*, 3:287–310, 1966. doi: 10.1137/0703023
38. A.J. Geurts. A contribution to the theory of condition. *Numer. Math.*, 39:85–96, 1982. doi: 10.1007/BF01399313
39. I. Gohberg and I. Koltracht. Mixed, componentwise, and structured condition numbers. *SIAM J. Matrix Anal. Appl.*, 14:688–704, 1993. doi: 10.1137/0614049

40. W. Wang and Y. Wei. Mixed and component-wise condition numbers for matrix decompositions. *Theoret. Comput. Sci.*, 681:199–216, 2017. doi: 10.1016/j.tcs.2017.03.034
41. W.W. Hager. Condition estimates. *SIAM J. Sci. Statist. Comput.*, 5:311–316, 1984. doi: 10.1137/0905023
42. N.J. Higham. *Accuracy and Stability of Numerical Algorithms*. SIAM, Philadelphia, PA, second edition, 2002. ISBN: 978-0-898715-21-7
43. C. Kenney and A.J. Laub. Condition estimates for matrix functions. *SIAM J. Matrix Anal. Appl.*, 10:191–209, 1989. doi: 10.1137/0610014
44. R. Mathias. Condition estimation for matrix functions via the Schur decomposition. *SIAM J. Matrix Anal. Appl.*, 16:565–578, 1995. doi: 10.1137/S0895479893244389
45. G.W. Stewart. *Matrix Algorithms*, volume I: Basic Decompositions. SIAM, Philadelphia, PA, 1998. ISBN: 0-89871-414-1
46. G.H. Golub and C.F. Van Loan. *Matrix Computations*. The John Hopkins University Press, Baltimore, MD, fourth edition, 2013. ISBN: 978-1-4214-0794-4
47. Å. Björck. *Numerical Methods in Matrix Computations*. Springer, New York, 2015. doi: 10.1007/978-3-319-05089-8
48. E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D.C. Sorensen. *LAPACK Users' Guide*. SIAM, Philadelphia, PA, third edition, 1999. ISBN: 0-89871-447-8
49. V.B. Lidskii. Perturbation theory of non-conjugate operators. *USSR Computational Mathematics and Mathematical Physics*, 6:73–85, 1966. doi: 10.1016/0041-5553(66)90033-4
50. J. Moro, J.V. Burke, and M.L. Overton. On the Lidskii–Vishik–Lyusternik perturbation theory for eigenvalues of matrices with arbitrary Jordan structure. *SIAM J. Matrix Anal. Appl.*, 18:793–817, 1997. doi: 10.1137/S0895479895294666
51. J.H. Wilkinson. *The Algebraic Eigenvalue Problem*. Clarendon Press, Oxford, UK, 1965. ISBN: 978-0-19-853418-1
52. G.W. Stewart. *Matrix Algorithms*, volume II: Eigensystems. SIAM, Philadelphia, PA, 2001. ISBN: 0-89871-503-2
53. J.W. Demmel. *Applied Numerical Linear Algebra*. SIAM, Philadelphia, PA, 1997. ISBN: 978-0-898713-89-3
54. R. Bhatia. *Perturbation Bounds for Matrix Eigenvalues*. Classics in Applied Mathematics. SIAM, Philadelphia, PA, 2007. doi: 10.1137/1.9780898719079
55. G.W. Stewart and J.-G. Sun. *Matrix Perturbation Theory*. Academic Press, San Diego, CA, 1990. ISBN: 978-0126702309
56. R. Li. Matrix perturbation theory. In L. Hogben, editor, *Handbook of Linear Algebra*, Discrete Math. Appl., pages (21–1)–(21–20). CRC Press, Boca Raton, FL, second edition, 2014. ISBN: 978-1-4665-0729-6
57. A. Greenbaum, R.-C. Li, and M.L. Overton. First-order perturbation theory for eigenvalues and eigenvectors. *SIAM Rev.*, 62:463–482, 2020. doi: 10.1137/19M124784X
58. L. Elsner. On the variation of the spectra of matrices. *Linear Algebra Appl.*, 47:127–138, 1982. doi: 10.1016/0024-3795(82)90230-0
59. L. Elsner. An optimal bound for the spectral variation of two matrices. *Linear Algebra Appl.*, 71:77–80, 1985. doi: 10.1016/0024-3795(85)90236-8
60. L. Elsner. Perturbation theorems for the matrix eigenvalue problem. *Port. Math.*, 43:69–76, 1986.
61. P. Henrici. Bounds for iterates, inverses, spectral variation and fields of values for non-normal matrices. *Numer. Math.*, 4:24–40, 1962. doi: 10.1007/BF01386294
62. Y. Nakatsukasa. *Algorithms and Perturbation Theory for Matrix Eigenvalue Problems and the Singular Value Decomposition*. PhD thesis, Office of Graduate Studies, University of California, Davis, CA, 2011.
63. J.-G. Sun. Stability and accuracy: Perturbation analysis of algebraic eigenproblems. Technical Report 98.07, Department of Computing Science, Umeå University, Umeå, Sweden, 1998. (Revised version, 2002).
64. J.H. Wilkinson. Sensitivity of eigenvalues. *Util. Math.*, 25:5–76, 1984.
65. J.H. Wilkinson. Sensitivity of eigenvalues, part II. *Util. Math.*, 30:243–286, 1986.
66. Z. Bai, J. Demmel, and A. Mckenney. On computing condition numbers for the nonsymmetric eigenproblem. *ACM Trans. Math. Software*, 19:202–223, 1993. doi: 10.1145/152613.152617
67. S.P. Chen, R. Feldman, and B.N. Parlett. Algorithm 517. A program for computing the condition numbers of matrix eigenvalues without computing eigenvectors. *ACM Trans. Math. Software*, 3:186–203, 1977. doi: 10.1145/355732.355741
68. M. Karow, D. Kressner, and F. Tisseur. Structured eigenvalue condition numbers. *SIAM J. Matrix Anal. Appl.*, 28:1052–1068, 2006. doi: 10.1137/050628519

69. R.M. Lin, Z. Wang, and M.K. Lim. A practical algorithm for the efficient computation of eigenvector sensitivities. *Computer Methods in Applied Mechanics and Engineering*, 130:355–367, 1996. doi: 10.1016/0045-7825(95)00929-9
70. J.-g. Sun. On condition numbers of a nondefective multiple eigenvalue. *Numer. Math.*, 61:265–275, 1992. doi: 10.1007/BF01385508
71. C. Van Loan. On estimating the condition of eigenvalues and eigenvectors. *Linear Algebra Appl.*, 88–89:715–732, 1987. doi: 10.1016/0024-3795(87)90131-5
72. S. Smale. The fundamental theorem of algebra and complexity theory. *AMS Bull. (New Series)*, 4:1–35, 1981. doi: 10.1090/S0273-0979-1981-14858-8
73. S. Smale. On the efficiency of algorithms of analysis. *AMS Bull. (New Series)*, 13:87–121, 1985. doi: 10.1090/S0273-0979-1985-15391-1
74. S. Smale. Some remarks on the foundations of numerical analysis. *SIAM Rev.*, 32:211–220, 1990. doi: 10.1137/1032043
75. J. Renegar. On the efficiency of Newton’s method in approximating all zeros of a system of complex polynomials. *Math. Oper. Res.*, 12:121–148, 1987. doi: 10.1287/moor.12.1.121
76. J. Renegar. Is it possible to know a problem instance is ill-posed? *J. Complexity*, 10:1–56, 1994. doi: 10.1287/moor.12.1.121
77. J.W. Demmel. On condition numbers and the distance to the nearest ill-posed problem. *Numer. Math.*, 51:251–289, 1987. doi: 10.1007/BF01400115
78. J.W. Demmel. The probability that a numerical analysis problem is difficult. *Math. Comp.*, 50:449–480, 1988. doi: 10.2307/2008617
79. J.W. Demmel. Steve Smale and the geometry of ill-conditioning. In M.W. Hirsch, J.E. Marsden, and M. Shub, editors, *From Topology to Computation: Proceedings of the Smalefest*, pages 305–316. Springer-Verlag, New York, 1993. doi: 10.1007/978-1-4612-2740-3
80. D.A. Klain and G.-C. Rota. *Introduction to Geometric Probability*. Lezioni Lincee. Cambridge University Press, Cambridge, UK, 1997. ISBN: 978-0-521-59362-5
81. L.A. Santaló. *Integral Geometry and Geometric Probability*. Cambridge University Press, New York, second edition, 2004. ISBN: 978-0-521-52344-8
82. H. Solomon. *Geometric Probability*. SIAM, Philadelphia, PA, 1978. ISBN: 978-0898710250
83. P. Bürgisser and F. Cucker. *Condition. The Geometry of Numerical Algorithms*. Springer, Heidelberg, 2013. ISBN: 978-3-642-38895-8
84. R. O. Akinola, M. A. Freitag, and A. Spence. The calculation of the distance to a nearby defective matrix. *Numer. Linear Algebra Appl.*, 21:403–404, 2013. doi: 10.1002/nla.1888
85. G. Armentia, J.-M. Gracia, and F.-E. Velasco. Nearest matrix with a prescribed eigenvalue of bounded multiplicities. *Linear Algebra Appl.*, 592:188–209, 2020. doi: 10.1016/j.laa.2020.01.026
86. A. Kalinina, A. Uteshev, M. Goncharova, and E. Lezhnina. On the distance to the nearest defective matrix. In F. Boulier, M. England, I. Kotsireas, T.M. Sadykov, and E.V. Vorozhtsov, editors, *Computer Algebra in Scientific Computing. CASC 2023.*, volume 14139 of *Lecture Notes in Computer Science*, pages 255–271, Cham, Switzerland, 2023. Springer Nature. doi: 10.1007/978-3-031-41724-5_14
87. R.A. Lippert. Fixing multiple eigenvalues by a minimal perturbation. *Linear Algebra Appl.*, 432:1785–1817, 2010. doi: 10.1016/j.laa.2009.11.030
88. A.N. Malyshev. A formula for the 2-norm distance from a matrix to the set of matrices with multiple eigenvalues. *Numer. Math.*, 83:443–454, 1999. doi: 10.1007/s002119900070
89. C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 3:211–218, 1936. doi: 10.1007/BF02288367
90. W. Kahan. Conserving confluence curbs ill-condition. Technical Report AD-766 916, University of California, Berkely, CA, August 1972.
91. A. Ruhe. Properties of a matrix with a very ill-conditioned eigenproblem. *Numer. Math.*, 15:57–60, 1970. doi:10.1007/BF02165660
92. J.H. Wilkinson. Note on matrices with a very ill-conditioned eigenproblem. *Numer. Math.*, 19:176–178, 1972. doi: 10.1007/BF01402528
93. D.J. Smith and M.K. Vamanamurthy. How small is a unit ball. *Mathematics Magazine*, 62:101–107, 1989. doi = 10.1080/0025570X.1989.11977419
94. H. Weyl. On the volume of tubes. *Amer. J. Math.*, 61:461–472, 1939. doi: 10.2307/2371513

95. R. Wongkew. Volumes of tubular neighbourhoods of real algebraic varieties. *Pacific Journal of Mathematics*, 159:177–184, 1993. doi: 10.2140/pjm.1993.159.177
96. M. Lotz. On the volume of tubular neighborhoods of real algebraic varieties. *Poc. Amer. Math. Soc.*, 143:1875–1889, 2014. doi: 10.1090/S0002-9939-2014-12397-5
97. A. Edelman. On the distribution of the scaled condition number. *Math. Comp.*, 58: 185–190. 1992. doi: 10.1090/S0025-5718-1992-1106966-2
98. P. Bürgisser, F. Cucker, and M. Lotz. The probability that a slightly perturbed numerical analysis problem is difficult. *Math. Comp.*, 77:1559–1583, 2008. doi: 10.1090/S0025-5718-08-02060-7
99. G. Frobenius. Theorie der linearen Formen mit ganzen Coefficienten. *Journal für die reine und angewandte Mathematik*, 86:146–208, 1879. doi: 10.3931/e-rara-18818
100. V.I. Arnold. On matrices depending on parameters. *Russian Math. Surveys*, 26:29–43, 1971. doi: 10.1070/rm1971v026n02abeh003827
101. J.W. Demmel and A. Edelman. The dimension of matrices (matrix pencils) with given Jordan (Kronecker) canonical form. *Linear Algebra Appl.*, 230:61–87, 1995. doi: 10.1016/0024-3795(93)00362-4
102. P.C. Hansen. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. SIAM, Philadelphia, PA, 1998. ISBN: 978-0-898714-03-6
103. G.H. Golub, V. Klementa, and G.W. Stewart. Rank degeneracy and least squares problems. Technical Report TR-456, Department of Computer Science, University of Maryland, College Park, MD, 1976.
104. G.W. Stewart. Rank degeneracy. *SIAM J. Sci. Stat. Comput.*, 5:403–413, 1984. doi: 10.1137/0905030
105. G.W. Stewart. On the early history of the singular value decomposition. *SIAM Rev.*, 35:551–566, 1993. doi: 10.1137/1035134
106. J. Demmel and B. Kågström. The generalized Schur decomposition of an arbitrary pencil $A - \lambda B$: Robust software with error bounds and applications. Part I: Theory and algorithms. *ACM Trans. Math. Software*, 19:160–174, 1993. doi: 10.1145/152613.152615
107. J. Demmel and B. Kågström. The generalized Schur decomposition of an arbitrary pencil $A - \lambda B$: Robust software with error bounds and applications. Part II: Software and applications. *ACM Trans. Math. Software*, 19:175–201, 1993. doi: 10.1145/152613.152616
108. B. Kågström and P. Wiberg. Extracting partial canonical structure for large scale eigenvalue problems. *Numer. Algorithms*, 24:195–237, 2000. doi: 10.1023/A:1019153528915
109. M.W. Hirsh and S. Smale. *Differential Equations, Dynamical Systems, and Linear Algebra*. Academic Press, San Diego, CA, 1974. ISBN: 0-12-349550-4
110. J. Palis and W. de Melo. *Geometric Theory of Dynamical Systems: An Introduction*. Springer-Verlag, New York, 1982. ISBN: 978-1-4612-5705-9
111. S. Wiggins. *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. Texts in Applied Mathematics. Springer, New York, second edition, 2003. ISBN: 0-387-00177-8-4
112. D.K. Arrowsmith and C.M. Place. *An Introduction to Dynamical Systems*. Cambridge University Press, Cambridge, UK, 1990. ISBN: 978-0-521-31650-7
113. S.-N. Chow, C. Li, and D. Wang. *Normal Forms and Bifurcation of Planar Vector Fields*. Cambridge University Press, London, UK, 1994. ISBN: 978-0-521-37226-8
114. A. Edelman, E. Elmroth, and B. Kågström. A geometric approach to perturbation theory of matrices and matrix pencils. part i: Versal deformations. *SIAM J. Matrix Anal. Appl.*, 18:653–692, 1997. doi: 10.1137/S0895479895284634
115. A. Edelman, E. Elmroth, and B. Kågström. A geometric approach to perturbation theory of matrices and matrix pencils. part ii: A stratification-enhanced staircase algorithm. *SIAM J. Matrix Anal. Appl.*, 20:667–699, 1999. doi: 10.1137/S0895479896310184
116. R. Gilmore. *Catastrophe Theory for Scientists and Engineers*. Dover Publications, New York, 1993. ISBN: 0-486-67539-4
117. L. Stolovitch. On the computation of a versal family of matrices. *Numer. Algorithms*, 4:25–46, 1993. doi: 10.1007/BF02142739
118. A.A. Mailybaev. Computation of multiple eigenvalues and generalized eigenvectors for matrices dependent on parameters. *Numer. Linear Algebra Appl.*, 13:419–436, 2006. doi: 10.1002/nla.471
119. A. Dmytryshyn. Versal deformations: A tool of linear algebra. *Cornell University Library, ArXiv e-prints in Representation Theory [math.RT]* 2312.14910v1, 1–15, 2023. doi: 10.48550/arXiv.2312.14910
120. M.W. Hirsh, S. Smale, and R.L. Devaney. *Differential Equations, Dynamical Systems, and an Introduction to Chaos*. Elsevier, Amsterdam, 2013. ISBN: 978-0-12-382010-5

121. B. Kågström and A. Ruhe. An algorithm for numerical computation of the Jordan normal form of a complex matrix. *ACM Trans. Math. Software*, 6:398–419, 1980. doi: 10.1145/355900.355912
122. B. Kågström and A. Ruhe. Algorithm 560: JNF, an algorithm for numerical computation of the Jordan normal form of a complex matrix. *ACM Trans. Math. Software*, 6:437–443, 1980. doi: 10.1145/355900.355917
123. Z. Zeng. Computing multiple roots of inexact polynomials. *Math. Comp.*, 74:869–903. doi: 10.1090/S0025-5718-04-01692-8.
124. Z. Zeng. Geometric modeling and regularization of algebraic problems. Cornell University Library, ArXiv e-prints in Numerical Analysis [math.NA] 2102.08830v1, 2021. doi: 10.48550/arXiv.2102.08330
125. Z. Zeng and T.-Y. Li. A numerical method for computing the Jordan canonical form. Cornell University Library, ArXiv e-prints in Numerical Analysis [math.NA] 2103.0206v1, 2021. doi: 10.48550/arXiv.2103.02086
126. V.N. Kublanovskaya. On a method of solving the complete eigenvalue problem for a degenerate matrix. *USSR Computational Mathematics and Mathematical Physics*, 6:1–14, 1966. doi: 10.1016/0041-5553(66)90001-2
127. A. Ruhe. An algorithm for numerical determination of the structure of a general matrix. *BIT Numerical Mathematics*, 10:196–216, 1970. doi: 10.1007/BF01936867
128. P. Petkov. *The Numerical Jordan Form*. World Scientific, Singapore, 2024. ISBN: 978-981-12-8644-5
129. V.I. Arnol'd. Lectures on bifurcations in versal families. *Russian Math. Surveys*, 27:54–123, 1972. doi: 10.1070/rm1972v027n05abeh001385
130. A.A. Mailybaev. Transformation to versal deformations of matrices. *Linear Algebra Appl.*, 337:87–108, 2001. doi: 10.1016/S0024-3795(01)00346-9
131. H. Poincaré. Sur l'équilibre d'une masse fluide animé d'un mouvement de rotation. *Acta Math.*, 7:259–380, 1885. doi: 10.1007/BF02402204
132. P. Glendinning. Bifurcation theory. In N.J. Higham, editor, *The Princeton Companion to Applied Mathematics*, pages 393–402. Princeton University Press, Princeton, NJ, 2015. ISBN: 978-0-691-15039-0
133. J.D. Crawford. Introduction to bifurcation theory. *Rev. Mod. Phys.*, 63:991–1037, 1991. doi: 10.1103/RevModPhys.63.991
134. Y.A. Kuznetsov. *Elements of Applied Bifurcation Theory*. Springer, New York, fourth edition, 2004. doi: 10.1007/978-3-031-22007-4
135. A. P. Seyranian and A. A. Mailybaev. *Multiparameter Stability Theory with Mechanical Applications*. World Scientific, Singapore, 2003. ISBN: 981-238-406-5

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.