Modern Birkhäuser Classics

V.I. Arnold
S.M. Gusein-Zade
A.N. Varchenko

Singularities of Differentiable Maps, Volume 1

Classification of Critical Points, Caustics and Wave Fronts

Birkhäuser

Modern Birkhäuser Classics

Many of the original research and survey monographs, as well as textbooks, in pure and applied mathematics published by Birkhäuser in recent decades have been groundbreaking and have come to be regarded as foundational to the subject. Through the MBC Series, a select number of these modern classics, entirely uncorrected, are being re-released in paperback (and as eBooks) to ensure that these treasures remain accessible to new generations of students, scholars, and researchers.

Singularities of Differentiable Maps

Volume 1

Classification of Critical Points, Caustics and Wave Fronts

V.I. Arnold
S.M. Gusein-Zade
A.N. Varchenko

Reprint of the 1985 Edition

Birkhäuser

V.I. Arnold (deceased)

A.N. Varchenko
Department of Mathematics
The University of North Carolina
at Chapel Hill
Chapel Hill, NC, USA

S.M. Gusein-Zade
Department of Mechanics and Mathematics
Moscow State University
Moscow, Russia

Originally published as Vol. 82 in the series *Monographs in Mathematics*

ISBN 978-0-8176-8339-9 ISBN 978-0-8176-8340-5 (eBook)
DOI 10.1007/978-0-8176-8340-5
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012938547

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.birkhauser-science.com)

V. I. Arnold
S. M. Gusein-Zade
A. N. Varchenko

Singularities of Differentiable Maps

Volume I

The Classification of
Critical Points
Caustics and Wave Fronts

Under the Editorship of V. I. Arnold
Translated by Ian Porteous
Based on a Previous Translation by Mark Reynolds

1985

Birkhäuser
Boston · Basel · Stuttgart

Originally published as
Osobennosty differentsiiruemykh otobrazhenii
by Science, Moscow, 1982.

Library of Congress Cataloging in Publication Data

Arnol'd, V. I. (Vladimir Igorevich), 1937–
Singularities of differentiable maps.

(Monographs in mathematics ; 82-)
Translation of: Osobennosti differentsiruemykh otobrazheniĭ.
Bibliography: p.
Includes index.
1. Differentiable mappings. 2. Singularities (Mathematics) I. Varchenko, A. N. (Aleksandr Nikolaevich) II. Guseĭn-Zade, S. M. (Sabir Medzhidovich) III. Title. IV. Series: Monographs in mathematics ; 82, etc.
QA614.58.A7513 1985 514'.72 84-12134
ISBN 0-8176-3187-9

CIP-Kurztitelaufnahme der Deutschen Bibliothek

Arnol'd, Vladimir I.:
Singularities of differentiable maps / V. I. Arnold ; A. N. Varchenko ; S. M. Gusein-Zade. – Boston ; Basel ; Stuttgart : Birkhäuser
Einheitssacht.: Osobennosti differenciruemych otobraženij «engl.»

NE: Varčenko, Aleksandr N.:; Gusejn-Zade, Sabir M.:

Vol. 1. The classification of critical points, caustics and wave fronts / under the editorship of V. I. Arnold. – 1985.

(Monographs in mathematics ; Vol. 82)
ISBN 3-7643-3187-9 (Basel ...)
ISBN 0-8176-3187-9 (Boston)

NE: GT

Printed in Switzerland by Birkhäuser AG, Graphische Unternehmen, Basel
ISBN 0-8176-3187-9
ISBN 3-7643-3187-9

The theory of singularities of differentiable maps is a rapidly developing area of contemporary mathematics, being a grandiose generalisation of the study of functions at maxima and minima and having numerous applications in mathematics, the natural sciences and technology (as in the so-called theory of bifurcations and catastrophes). Chapters of the book are devoted to the theory and stability of smooth maps, critical points of smooth functions, and the singularities of caustics and wave fronts in geometrical optics.

The book is the first volume of a large monograph envisaged by the authors. In the second volume will be placed the algebro-topological aspects of the theory.

The book is intended for mathematicians, from second year students up to research workers, and also for all users of the theory of singularities in mechanics, physics, technology and other sciences.

VLADIMIR IGOREVICH ARNOLD
ALEKSANDR NIKOLAEVICH VARCHENKO
SABIR MEDZHIDOVICH GUSEIN-ZADE

Introduction to the English Edition

Singularity theory is still in the state of very rapid development and many new results appeared after the Russian edition of this book. The reader should consult the two volumes of the Arcata singularities conference (Symposia in Pure Math., vol. 40, 1983.); the surveys "Singularities of ray systems", Russ. Math. Surveys (vol. 38, no. 2, 1983); "Singularities in the calculus of variations", Contemporary Problems of Math., (vol. 22, Moscow, 1983); Catastrophe Theory, Springer, 1983; and the Proceedings of the International Congress of Mathematics, Warsaw, 1983. The authors express their gratitude to R. and C. MacPherson and to I. and H. Porteous for their care with the English edition.

Foreword

... there is nothing so enthralling, so grandiose, nothing that stuns or captivates the human soul quite so much as a first course in a science. After the first five or six lectures one already holds the brightest hopes, already sees oneself as a seeker after truth. I too have wholeheartedly pursued science passionately, as one would a beloved woman. I was a slave, and sought no other sun in my life. Day and night I crammed myself, bending my back, ruining myself over my books; I wept when I beheld others exploiting science for personal gain. But I was not long enthralled. The truth is every science has a beginning, but never an end – they go on for ever like periodic fractions. Zoology, for example, has discovered thirty-five thousand forms of life ...

A. P. Chekhov. "On the road"

In this book a start is made to the "zoology" of the singularities of differentiable maps. This theory is a young branch of analysis which currently occupies a central place in mathematics; it is the crossroads of paths leading from very abstract corners of mathematics (such as algebraic and differential geometry and topology, Lie groups and algebras, complex manifolds, commutative algebra and the like) to the most applied areas (such as differential equations and dynamical systems, optimal control, the theory of bifurcations and catastrophes, short-wave and saddle-point asymptotics and geometrical and wave optics).

The main applications of the theory of singularities consist in the listing and detailed examination in each situation of a small collection of the most frequently encountered standard singularities, which are just those which occur for objects in general position: all more complicated singularities decompose into the simplest ones under a small perturbation of the object of study. We give rather complete lists, diagrams and determinators of these simplest singularities for a whole series of objects (functions, maps, varieties, bifurcations, caustics, wave fronts and the like) trying as far as possible to shorten for the reader the path from the beginnings of the theory to its applications. Accordingly our aim is to present the basic ideas, methods and results of the theory of singularities in such a way that the reader may as quickly as possible learn to apply the methods and results of the theory, without being delayed on the more fundamental theological parts.

A special effort has been made to ensure that the application of the main ideas and methods is not clogged up with technical details. The most fundamental and the simplest examples are studied in the greatest detail, while the presentation of the more specialised and difficult parts of the theory has the character of a review.

The reader is supposed to have only a small mathematical knowledge (skill in differentiating and some linear algebra and geometry)*. The authors have tried to present the material in such a way that the reader can omit parts that he finds difficult without significantly impairing his understanding of what follows.

At the present time the theory of singularities is rapidly developing (see, for example, the lists of unsolved problems in [19] and [27]) and we shall not attempt to cover all the many directions of current research (an incomplete bibliography of approximately 500 works is to be found in Poston and Stewart [148] and Brieskorn [41]).

The basis of this book is a series of special courses given at the Mechanics-Mathematics Faculty of Moscow State University during the years 1966–1978. In its preparation we have used lecture notes taken by V. A. Vasil'ev, E. E. Landis and A. G. Hovansky; A. G. Hovansky has written Chapter 5. The authors wish to thank these people for their assistance, as well as to thank the participants of a seminar on the theory of singularities, whose assistance they have made great use of, especially A. G. Kushnirenko, E. I. Korkina and V. I. Matov.

The complex-analytic and algebro-geometric aspects of singularity theory (monodromy, intersection theory, asymptotics of integrals and mixed Hodge structures) will be discussed in a second volume "Singularities of differentiable maps, Algebro-topological aspects" in course of preparation.

Yasenevo, March 1979

*For the benefit of the reader we recall the following terminology:
(1) A *manifold* is an n-dimensional generalisation of a curve or surface, while a *map* is the analogous generalisation of function. A *diffeomorphism* is an invertible map, it and its inverse both being differentiable.
(2) A *transformation* of a set is an invertible map of the set to itself. A *group of transformations* of a set is a set of transformations containing together with each transformation its inverse and together with every two transformations their composite. A *group* results from axiomatising the properties of a group of transformations.
(3) An *algebra* results from axiomatising the properties of the set of all functions on a set (the elements of an algebra, just like functions, may be added or multiplied together or may be multiplied by numbers, and moreover the usual rules of associativity, distributivity and commutativity hold. In an algebra there is a distinguished element 1 with $1f \equiv f$).
(4) A *module over an algebra* is the result of axiomatising the properties of the set of all vector fields on a manifold (the elements of a module may be added together and multiplied by elements of the algebra).
(5) An *ideal* in an algebra is a subset of the algebra that is also a module over the algebra. Example: in the algebra of all the functions on a manifold the functions taking the value zero on a given submanifold form an ideal.

Table of Contents

Part I
Basic Concepts

The theory of singularities of smooth maps is a wide-ranging generalisation of the theory of the maxima and minima of functions of one variable. Therefore the singularities that we shall be discussing are not connected with discontinuities and poles, but with the vanishing of certain derivatives and Jacobians.

In this Part the basic concepts in the theory of the singularities of differentiable maps will be introduced: singular points, their local algebras and other invariants; concepts concerned with stability will be defined and a start made to the classification of singularities.

1. The simplest examples

Here we describe the classification due to H. Whitney of the singularities of smooth maps of spaces of small dimensions.

1.1 Critical points of functions

A point x is said to be a *critical point* of a function f, if at that point the derivative of f is zero.

Example: Let $f: \mathbb{R} \to \mathbb{R}$ be the function, given by the formula $y = x^2$. Then the point 0 is a critical point of the function.

The critical points of functions are divided into generic or nondegenerate critical points and degenerate critical points.

Definition: A critical point of a smooth function is said to be *nondegenerate* if the second differential of the function at that point is a nondegenerate quadratic form.

Example: The critical point 0 of the function $y = x^2$ is nondegenerate, while the critical point 0 of the function $y = x^3$ is degenerate (Fig. 1).

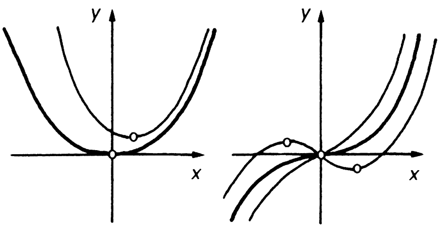

Fig. 1.

Consider an arbitrary smooth function, close (along with its derivatives) to the function $y = x^2$. It is clear that near zero this function will have a critical

V.I. Arnold et al., *Singularities of Differentiable Maps, Volume 1*, Modern Birkhäuser Classics,
DOI 10.1007/978-0-8176-8340-5_1, © Springer Science+Business Media New York 2012

point, similar to the critical point of $y = x^2$. The critical point of $y = x^2$ is stable in the sense that under small perturbations of the function it does not vanish, but simply shifts slightly.

The degenerate critical point of the function $y = x^3$ behaves completely differently under small perturbations.

Example: Consider the family of functions of one variable $y = x^3 + \varepsilon x$. For small ε the functions of this family can be considered as small perturbations of $y = x^3$. We see that under this perturbation the degenerate critical point either vanishes (for $\varepsilon > 0$) or decomposes into two nondegenerate critical points at a distance of order $\sqrt{|\varepsilon|}$ from it (for $\varepsilon < 0$).

Thus the critical point of $y = x^2$ is stable, while that of $y = x^3$ is unstable.

It is not difficult to analyse all possible situations for a function of one variable. We consider the space of all functions of interest to us*

We distinguish in this space the set of functions having degenerate critical points or having coincident values at different critical points (Fig. 2.). [In the case where the domain of definition is a line segment we consider an end-point to be critical, counting it as nondegenerate if the derivative there is nonzero.] It is not difficult to see that such "degenerate" functions form a thin set, to be precise a hypersurface or surface of codimension one "given by one equation" in our function space. We shall later give a precise meaning to this phrase and prove a corresponding theorem in a more general situation. This hypersurface divides our function space into parts, in each of which all the functions are "constructed in the same way": in each part the order relations between the values at successive critical points are the same for each function in the region.

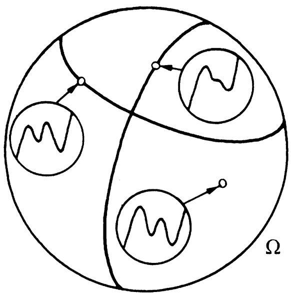

Fig. 2.

* This could be the space of all infinitely differentiable or sufficiently differentiable functions, or the space of analytic functions, or even the space of polynomials; it is convenient to suppose that the domain of definition of functions is compact, considering functions on a circle or line segment.

Functions having neither degenerate critical points nor multiple critical values are said to be *Morse functions.* Under a small perturbation a Morse function "preserves its form" and can be transformed to the original function by smooth changes of the independent and dependent variables x and y. In this sense a Morse function is stable. Thus in the case where the source and target spaces of the maps are one-dimensional the stable maps form an open everywhere dense set in the space of all maps. Moreoever, the stable maps admit a sufficiently clear description and classification, while the unstable maps, though constructed possibly in a much more complicated manner (the set of critical points can be an arbitrary closed set), are transformed into stable ones by a small perturbation: each complicated singularity breaks up into several nondegenerate stable singularities.

The study of maps to the line (Morse theory) provides a special case for which the goals of singularity theory can be attained. The results of Morse theory of interest to us can be stated as follows.

Theorem: (1) *The stable maps* $f: M^m \to \mathbb{R}^1$ *of a closed* manifold* M^m *to the line form an everywhere dense set in the space of smooth maps.*

(2) *For a map* f *to be stable it is necessary and sufficient that the following two conditions be satisfied:*

M_1. *The map* f *is stable at each point (in other words all the critical points of the function are nondegenerate).*

M_2. *All the critical values of the function are distinct.*

(3) *A map* $f: M^m \to \mathbb{R}^1$ *is stable at a point* x_0 *if and only if there are neighbourhoods of* $x_0 \in M^m$ *and* $y_0 = f(x_0) \in \mathbb{R}^1$ *in which coordinates* $x_1, \ldots, x_m; y$ *can be chosen so that the map can be written in one of the following* $m + 2$ *forms:*

$MI.\ y = x_1$

$MII_k.\ y = x_1^2 + \cdots + x_k^2 - x_{k+1}^2 - \cdots - x_m^2 \quad (k = 0, 1, \ldots, m).$

For a proof see, for example, [131].

The question arises as to whether such a result holds true in more dimensions, that is for maps $f: M^m \to N^n$ of manifolds of arbitrary dimensions m and n.

1.2 Critical points and critical values of smooth maps

Consider a differentiable map $f: M^m \to N^n$. First of all we must extend to this case the concept of critical point. The derivative of a map f at a point x is a

* Here and in the sequel a closed manifold will mean a compact manifold without boundary.

linear map of the tangent space of the source manifold at the point x to the tangent space of the target manifold at the point $f(x)$:

$$f_{*x}: T_x M^m \to T_{f(x)} N^n.$$

Example: Let M^2 be the surface of a sphere in three-dimensional space, N^2 the plane and f the projection of the sphere vertically on to the horizontal plane (Fig. 3).

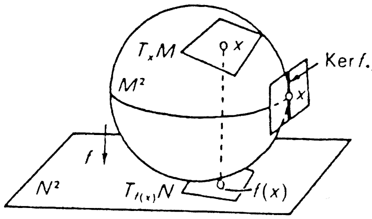

Fig. 3.

The linear map f_{*x} of the plane $T_x M$, tangent to the sphere at x, to the plane $T_{f(x)} N$, tangent to the horizontal plane, is a nondegenerate linear map if the point x does not belong to the horizontal equator of the sphere. If, however, x lies on the equator, then the tangent plane to the sphere at x contains a vertical line. In this case the projection operator has a nontrivial kernel (the subspace mapped to zero). The kernel of f_{*x} at points of the equator is one-dimensional. The rank of f_{*x} at these points is equal to 1.

We make now the more general

Definition: A point x of the manifold M is said to be a *critical point* of the smooth map $f: M \to N$ if the rank of the derivative

$$f_{*x}: T_x M \to T_{f(x)} N$$

at that point is less than the maximum possible value, that is less than the smaller of the dimensions of M and N:

$$\operatorname{rank} f_{*x} < \min(\dim M, \dim N).$$

Remark: Let $x_1, \ldots, x_m$ be local coordinates in a neighbourhood of the point x

in M and let $y_1, \ldots, y_n$ be local coordinates in a neighbourhood of the point $f(x)$ in N. In terms of these coordinates the map f is given by n smooth functions in m variables:

$$y_1 = f_1(x_1, \ldots, x_m), \ldots, y_n = f_n(x_1, \ldots, x_m).$$

The matrix $(\partial f_i/\partial x_j)$ is called the *Jacobian matrix* of the map. In these terms one can say that *the point x is critical if the rank of the Jacobian matrix there is not maximal.*

Example: For the projection of the sphere to the horizontal plane the critical points are the points of the horizontal equator. Off the equator the rank of the derivative is equal to 2, while at points x of the equator the rank of the operator f_{*x} falls to 1.

The image of a critical point is called a *critical value.*

Example: The critical values of the projection of the sphere to the plane form the circle of the apparent contour of the sphere.

1.3 Differentiable equivalence

There are several different ways of classifying smooth maps. Clearly the crudest classification is the topological one: two maps are said to be *topologically equivalent* if there are homeomorphisms (one-to-one continuous maps with continuous inverses) of the source and target manifolds, transforming the one map into the other. The functions $y = x^2$ and $y = x^4$ are topologically equivalent.

If $f_p: M_p \to N_p$, $p = 1, 2$ are two given maps then to say that they are topologically equivalent means that there exist homeomorphisms $h: M_1 \to M_2$ and $k: N_1 \to N_2$ such that $f_2 = kf_1h^{-1}$.

In other words a topological equivalence is a commutative diagram

$$\begin{array}{ccc} M_1 & \xrightarrow{f_1} & N_1 \\ h\downarrow & & \downarrow k \\ M_2 & \xrightarrow{f_2} & N_2 \end{array}$$

in which the vertical arrows are homeomorphisms.

For the purposes of analysis topological equivalence, as a rule, is too crude a concept. For example, the function $y = x^4$ with a degenerate unstable singularity is topologically equivalent to a stable one. Therefore in the theory of singularities a second concept is basic, the concept of differentiable equivalence.

Definition: A *differentiable equivalence* of differentiable maps $f_1 : M_1 \to N_1$ and $f_2 : M_2 \to N_2$ is a commutative diagram

$$\begin{array}{ccc} M_1 & \xrightarrow{f_1} & N_1 \\ h \downarrow & & \downarrow k \\ M_2 & \xrightarrow{f_2} & N_2 \end{array}$$

whose vertical arrows are diffeomorphisms (differentiable one-to-one maps whose inverses also are differentiable*).

Remark 1: In the language of local coordinates the map $y = f(x)$ is a set of functions, the diffeomorphism h is a change of the independent variables x and the diffeomorphism k is a change of the dependent variables y. From this point of view the question of differentiable equivalence is the question whether one can transform one map into the other by means of smooth changes of the independent and dependent variables.

Remark 2: The commutative diagram depicted above represents the identity

$$k(f_1(h^{-1}(x))) \equiv f_2(x).$$

In this formula h^{-1} lies to the *right* of f_1 and k to the *left*. Therefore the diffeomorphism h^{-1} of the source space (and the change of the independent variables x) is said to be a *right change*. Likewise the diffeomorphism k of the target space (and the change of the dependent variables y) is said to be a *left change*.

* Here and in the sequel unless there is explicit mention to the contrary the word *differentiable* or *smooth* means "continuously differentiable the necessary number of times", for example infinitely differentiable.

Remark 3: Yet another way of expressing the same thing consists in the following. Consider the set $\Omega(M, N)$ of all smooth maps from M to N. Consider the group Diff M of all diffeomorphisms of the source manifold M with itself and the group Diff N of all diffeomorphisms of the target manifold N with itself.

The direct product of groups

$$\text{Diff}\, M \times \text{Diff}\, N$$

consists of all pairs (h, k) of diffeomorphisms of the source space $(h: M \to M)$ and the target space $(k: N \to N)$.

The group Diff $M \times$ Diff N acts on the set $\Omega(M, N)$ in the following way: if $f \in \Omega(M, N)$, $h \in \text{Diff}\, M$, $k \in \text{Diff}\, N$, then $(h, k)f = k \circ f \circ h^{-1}$.

It is not difficult to verify that this is a genuine group action, that is that

$$(h_1 h_2, k_1 k_2)f = (h_1, k_1)((h_2, k_2)f).$$

This action is said to be *left-right* (the action of *Diff M* is said to be *right* and the action of Diff N to be *left*).*

In these terms we can reformulate the definition of differentiable equivalence as follows: *two maps from M to N are differentiably equivalent if and only if they belong to the same orbit of the left-right action* (Fig. 4).

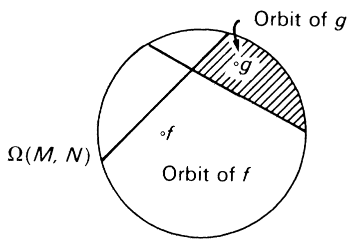

Fig. 4.

Example: The connected components of the set of all Morse functions (defined in section 1.1) are the orbits of the left-right action.

1.4 Stability

Consider a smooth map $f: M \to N$ of a closed manifold M to a manifold N.

* Not to be confused with left action in algebraic terminology.

Definition: A map f is said to be *differentiably stable* (or more precisely *left-right-differentiably stable*, or briefly simply *stable*), if every map sufficiently close to it† is differentiably equivalent to it.

In other words, f is stable, if its left-right orbit is open.

Example: The projection map of the sphere to the plane is stable. The map of the circle $\{x \bmod 2\pi\}$ to the line $\{y\}$, given by the formula $y = \sin 2x$, is unstable.

Remark: If one replaces differentiable equivalence in the preceding definition by topological then one obtains the definition of *topological stability*.

Example: The projection map of the sphere to the plane is topologically stable, as is any differentiably stable map. Topologically stable but differentiably unstable maps exist, but it is not easy to give an example (see [128]). The formula $y = \sin 2x$ defines a topologically unstable map from the circle to the line.

There exist also local versions of the concepts that we have introduced. For example, the singularity at zero of the function $y = x^2$ is stable, while the singularity at zero of the function $y = x^3$ is unstable. To give a formal definition of the stability of a map at a point we employ the following terminology.

Definition: A *map-germ* $M \to N$ at a point x of M is an equivalence class of maps $\varphi : U \to N$ (each of which is defined on some neighbourhood U of x in M, not necessarily the same for each); here two maps are regarded as equivalent if they coincide on some neighbourhood of the point x (this neighbourhood may well be smaller than the intersection of the neighbourhoods on which the two maps are defined). Two maps of the same class are said to have the *same germ at the point* x (Fig. 5).

In other words the germ of the map f at x is what remains of it when "its domain of definition is made infinitely small".

† Differing from f by a sufficiently small amount, provided that a sufficiently large number of derivatives are taken into account.

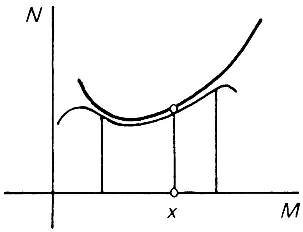

Fig. 5.

Definition: Two smooth map-germs are said to be (*left-right*, *differentiably*) *equivalent* if there are germs of diffeomorphisms of the source and target spaces transforming the first germ into the second (if the map-germ f_1 at x_1 is equivalent to the map-germ f_2 at x_2 then there exist a diffeomorphism-germ h at x_1 sending x_1 to x_2 and a diffeomorphism-germ k at $f_1(x_1)$ sending $f_1(x_1)$ to $f_2(x_2)$, such that $k(f_1(h^{-1}(x))) \equiv f_2(x)$ in a sufficiently small neighbourhood of x_2). The equivalence class of a germ at a critical point is said to be a *singularity*.

Definition: A smooth map-germ $f: M \to N$ at a point x of M (Fig. 6) is said to be (*left-right*, *differentiably*) *stable* if for a sufficiently small neighbourhood U of x there is a neighbourhood E of the map* f in $\Omega(M, N)$, such that for any map $\tilde{f}$ in E there is a point $\tilde{x}$ in U such that the germ of $\tilde{f}$ at $\tilde{x}$ is equivalent to the germ of f at x.

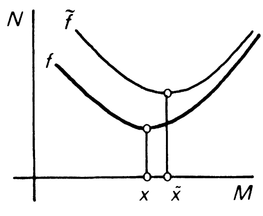

Fig. 6.

* A neighbourhood of a given map is the set of all maps, differing only slightly from the given one when derivatives up to a fixed order are taken into account. In the local situation of interest to us at the moment we can suppose that $M \subset \mathbb{R}^m$ and $N \subset \mathbb{R}^n$ are regions of Euclidean spaces and that E is given by the inequalities $\|(\tilde{f}-f)\|_k < \varepsilon$, where $\|g\|_k = \sup_{|\alpha| \leqslant k} |\partial^\alpha g/\partial x^\alpha|$.

It is not difficult to verify that stability at a point is a property of the germ and not of the map: the property is not destroyed by changes in f that leave alone some neighbourhood of x.

The *topological equivalence of germs* and the *topological stability of germs* are defined in similar fashion.

Example: The map-germs $y = x^2$ and $y = x^4$ at the point 0 of the real line are topologically equivalent. The germ $y = x^2$ is topologically (and even differentiably) stable at zero. The germ $y = x^4$ is differentiably (and even topologically) unstable at zero.

1.5 The stability of maps from two-dimensional manifolds to two-dimensional manifolds.

We begin with an example that we looked at earlier – the projection map of the sphere to the plane (Fig. 3). The singularities of the projection are the points of the equator of the sphere. It is not difficult to verify that the map-germ at each point of the equator is stable.

It is difficult to imagine how other stable singularities of maps of two-dimensional manifolds to two-dimensional manifolds can exist. In fact what we see when we look at the smooth surfaces of three-dimensional bodies are their apparent contours, consisting of the critical values of the projection map of the surface to the retina of the eye. It usually seems to us that these apparent contours consist of smooth curves. However, if we look about us more attentatively (for example, at the faces of the people around us) then we may discern alongside singularities like the singularities of the projection of the equator of the sphere singularities of yet another type. These singularities were discovered by H. Whitney, who in 1955 [192] gave a complete description of the singularities of a generic map from a two-dimensional manifold to a two-dimensional manifold. This work of Whitney is the foundation of singularity theory whose date of birth can therefore be taken to be 1955.

Whitney established that *every smooth map of a compact two-dimensional manifold to a two-dimensional manifold can be approximated arbitrarily closely (for any number of derivatives) by a stable map.*

Moreover he studied the structure of stable maps. At each point the germ of such a map is stable. Whitney described all stable map-germs of two-dimensional manifolds (there are three of them, up to differentiable equivalence). Finally he proved that a map of a compact two-dimensional manifold to a two-dimensional manifold is stable if its germ at each point is stable and the critical values are disposed "in a general manner" (this being a generalisation of the condition of non-coincidence of critical values at different critical points for Morse functions, see Section 1.1).

Whitney's Theorem: *A map of a two-dimensional manifold to a two-dimensional manifold is stable at a point if and only if the map can be described with respect to local coordinates* (x_1, x_2) *in the source and* (y_1, y_2) *in the target in one of the three forms:*

WI $y_1 = x_1$, $y_2 = x_2$ (*a regular point*);
WII $y_1 = x_1^2$, $y_2 = x_2$ (*a fold*);
WIII $y_1 = x_1^3 + x_1x_2$, $y_2 = x_2$ (*a pleat*)

(*the point under consideration has the coordinates* $x_1 = x_2 = 0$).

In other words, every stable map-germ of a two-dimensional manifold to a two-dimensional manifold is differentiably equivalent to one of the three map-germs at zero of the given list.

The first of the given germs is a diffeomorphism-germ. Every smooth map of a two-dimensional manifold reduces to this form in the neighbourhood of a noncritical point.

A singularity of a map of the second kind is said to be a *fold*. This map of the plane to the plane can be considered as a family of maps of the line to the line ($y_1 = x_1^2$), depending trivially on one parameter ($y_2 = x_2$).

Example: The projection of the sphere to the horizontal plane has a singularity of fold type on the equator, as one may easily verify, by choosing suitable local coordinates (x_2 and y_2 as longitude and x_1 as latitude, Fig. 7).

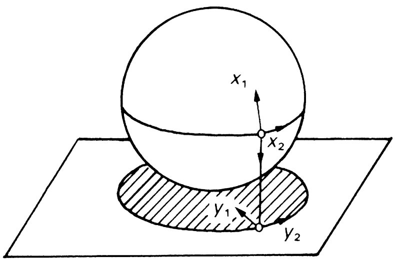

Fig. 7.

1.6 The Whitney pleat

The *Whitney pleat*, frequently called the *Whitney cusp*, is the third stable singularity of the above list, that is the singularity of the map

$$y_1 = x_1^3 + x_1 x_2, \; y_2 = x_2$$

at zero. To see this singularity clearly we realise it as a singularity of the vertical projection of a smooth surface in three-dimensional space to the horizontal plane.

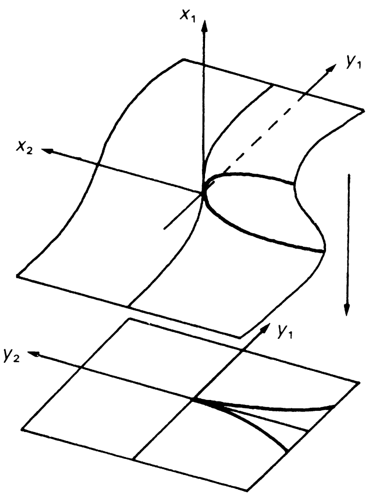

Fig. 8.

With this in view, consider the graph of the function $y_1 = x_1^3 + x_1 x_2$ in three dimensional space with coordinates (x_1, x_2, y_1) (Fig. 8). This graph is diffeomorphic to the plane (every graph of a smooth map is diffeomorphic to its domain): as coordinates on the graph we may take x_1 and x_2. Consider the intersection of the graph with the vertical plane $x_2 =$ const. For a fixed value of x_2 the equation $y_1 = x_1^3 + x_1 x_2$ determines a cubic curve, lying in the vertical plane.

If $x_2 > 0$ then along the corresponding cubic curve y_1 increases monotonically with x_1. If however $x_2 < 0$ then y_1 has two critical points – a local maximum and a local minimum.

Consider now the projection of our graph to the horizontal plane, $(x_1, x_2, y_1) \mapsto (x_2, y_1)$. The resulting smooth map of the surface to the plane has a singularity of pleat type at the origin. In fact consider the following systems of coordinates: (x_1, x_2) on the graph and $(y_1, y_2 = x_2)$ on the horizontal plane. With respect to these coordinates the projection is described exactly by the formulas of the Whitney pleat. Whitney's theorem asserts that the singularity is stable. In particular under a small perturbation of our surface in three-dimensional space a surface is formed whose projection to the horizontal plane has a singularity of the same type at some point near to the origin.

Problem: Find the critical points of the Whitney map

$$y_1 = x_1^3 + x_1 x_2, \; y_2 = x_2.$$

Solution: The Jacobian matrix has the form

$$\left(\frac{\partial y}{\partial x}\right) = \begin{pmatrix} 3x_1^2 + x_2 & x_1 \\ 0 & 1 \end{pmatrix}.$$

At critical points the rank of this matrix is less than two, that is its determinant is equal to zero: $3x_1^2 + x_2 = 0$. Consequently the set of critical points is a smooth curve. In the (x_1, x_2) plane the equation $3x_1^2 + x_2 = 0$ determines a parabola (Fig. 9).

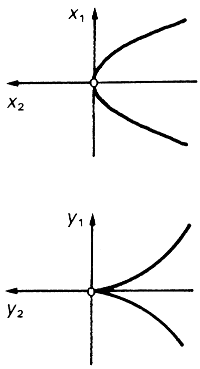

Fig. 9.

On our graph in three-dimensional space the critical points of the projection are those points where the tangent plane to the graph contains a vertical line. It is clear that these are just the critical points of the functions y_1 on the cubic curves described above. We conclude that all these points form a smooth curve on the graph (this is not obvious without calculation).

Problem: Find the critical values of the Whitney map.

Solution: At critical points $x_2 = -3x_1^2$. Substituting for x_2 its expression in terms of x_1, we obtain the parametric equations of the set of critical values:

$$y_1 = x_1^3 + x_1x_2 = -2x_1^3,\ y_2 = x_2 = -3x_1^2.$$

Thus the set of critical values is a semicubical parabola on the (y_1, y_2) plane. This curve has a singular point (a cusp, also called a point of regression) at the origin. It divides the plane into two parts. Under the Whitney map every point in the smaller part has three preimages while a point in the larger has only one.

Remark: If one looks down on the surface of the graph then one sees only the upper fold, ending at the pleat point, and it has the form of one half of a semicubical parabola. Most of the surfaces that we see are opaque surfaces. Therefore we usually see that the fold terminates at a pleat point, but we do not notice the cusp.

Example: Consider the surface of a torus in three-dimensional space as depicted in Fig. 10. Two pleat points are clearly seen. If the torus was transparent then we would see the picture drawn in Fig. 11, with four pleats. We seldom come across transparent tori. More frequently we come across the smooth surface of

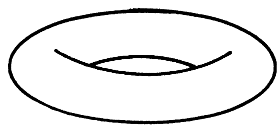

Fig. 10.

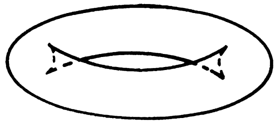

Fig. 11.

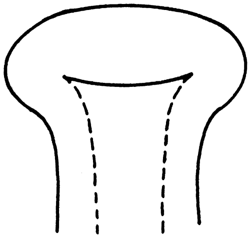

Fig. 12.

a glass bottle. If one looks at the neck it is easy to make out two pleats (Fig. 12). By moving the bottle one can convince oneself of their stability.

1.7 Catastrophes

Whitney's theorem asserts that folds and pleats are not destroyed by small perturbations, while all more complicated singularities break up under a small perturbation into folds and pleats and therefore do not occur for generic smooth maps of two-dimensional surfaces. One comes across smooth maps everywhere. By Whitney's theorem we must therefore everywhere come across fold contours and cusps of semicubical parabolas on them. This remarkable theorem has given birth to much speculation, associated principally with the names of R. Thom and C. Zeeman. They gave the name of *catastrophe theory* to the whole subject of applications of Whitney's theorem. Hundreds of papers have been published and are being published on catastrophe theory, being mainly concerned with applications. The usual form of argument in catastrophe theory is as follows. One considers a smooth surface in three-dimensional space about which usually almost nothing is known, together with a projection map to the plane. Since nothing is known about the surface it is supposed to be a surface in general position. In that case the singularities of the projection are folds and pleats. Already this alone leads to some information about the jumps or "catastrophes" which the objects of study can undergo.

We give below an example of such an "application". It is more or less borrowed from the work of Zeeman.

Let us characterise the creative personality (say of a scholar) by three parameters, which we call "technique", "enthusiasm" and "achievement". Clearly these parameters must be interdependent. In fact a surface is obtained in three-dimensional space with coordinates (T, E, A). We project this surface to the (T, E)

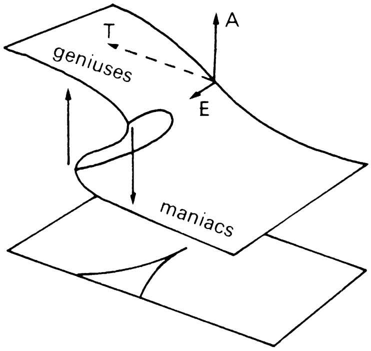

Fig. 13.

plane. For a surface in general position the singularities of the projection are folds and pleats. It is asserted that a pleat, placed as depicted in Fig. 13, satisfactorily describes the phenomena under investigation.

Consider, in fact, how under these hypotheses the achievement of a scholar varies with his technique and enthusiasm. If his enthusiasm is weak then his achievement will increase monotonically and rather slowly with his technique. If, however, his enthusiasm is sufficiently great then qualitatively new phenomena appear. In this case his achievement may increase with a jump as his technique increases. The area of high achievement which we reach in this way is indicated on our diagram by the word "geniuses".

On the other hand, an increase in enthusiasm not reinforced by a corresponding increase in technique leads to an area indicated on our diagram by the word "maniacs". It is noteworthy that the "catastrophes", the jumps from the state of "genius" to the state of "maniac" and back again, take place on different lines, so that with a sufficiently high degree of enthusiasm geniuses and maniacs may have the same amount of technique but differing amounts of achievement.

The defects in the model we have described are too obvious for it to be necessary for us to speak of them in more detail (see, for example, [148]).

There are serious applications of the theory of singularities, for example in the theory of elasticity, in optics (the singularities of caustics and wave fronts), in the theory of oscillating integrals (the method of stationary phase), and so on. We return to these applications after we have developed the appropriate techniques.