CHARACTERIZATION OF FUNCTIONS REPRESENTING TOPOGRAPHIC SURFACES

Gert W. WOLF
Department of Geography, University of Klagenfurt
Universitaetsstrasse 65, A-9010 Klagenfurt, Austria
e-mail: WOLF@EDVZ.UNI-KLAGENFURT.ADA.AT

ABSTRACT

During the past years it has become apparent that a general framework of spatial data management resting on formal methods is indispensable. One aspect of such an analytic framework is the adequate characterization of functions so that they may be regarded as abstract models of real topographic surfaces. The importance of a precise mathematical description like this results from the fact that the theoretical requirements of differentiability and continuity of the derivatives, which are commonly employed in practical applications, do not suffice for functions to represent realizable topographic surfaces. The reason for this failure is that continuously differentiable mappings may still be endowed with some pecularities which are extremely unlikely to appear in reality and thus prevent the functions from being suitable models for the topography of a given area. It will be demonstrated that a great many of these pecularities are due to structural instability - a phenomenon which can easily be explained by the presence or absence of degenerate critical points and saddle connections. Since it can be proved that any function possessing degenerate critical points may be approximated accurately enough by another one without such points, mappings of the latter type (so-called Morse functions) which have, in addition, no saddle connections should describe topographic surfaces in an appropriate way. The results arrived at in this paper, however, are valid not only for functions defined on the plane but also for mappings defined on differentiable manifolds and thus help to diminish the deficiency of theoretical knowledge concerning curved surfaces as has been complained recently.

1.INTRODUCTION

As a consequence of the numerous applications of computers in cartography during the past years it has become apparent that a general framework of spatial data management and analysis is indispensable. This realization has given rise to an increasing number of publications concerning the formal foundations of numerous cartographic concepts. The different approaches covered a wide portion of the field of cartography ranging from the development of analytic tools for cartographic generalization (e.g. WOLF 1988a,b, 1989, WEIBEL 1989) to the design of databases for geographic information systems (e.g. PEUCKER 1973, PEUCKER/CHRISMAN 1975, PEUQUET 1983, BOUDRIAULT 1987, SALGÉ/SCLAFER 1989).

The subject of the present paper is the formal analysis of another point of interest in computer cartography, namely the adequate characterization of functions so that they may be regarded as abstract models of real topographic surfaces. The importance of a formal characterization like this is derived above all from the following four facts: First of all, theoretical results obtained for functions describing topography hold also for functions describing phenomena like population density, accessibility, pollution, temperature, precipitation etc.[1]; secondly, topographic surfaces represent the underlying continuous model of DTMs whereby DTM may stand as abbreviation for 'digital terrain model' or 'discrete terrain model' respectively; thirdly, a great many of the results derived for mappings from $R^2 \to R$ are also true for real-valued mappings defined on curved surfaces - so-called differentiable manifolds. As it has been pointed out just recently especially this point deserves our special attention '(since) geographical data (are) distributed over the curved surface of the earth, a fact which is often forgotton ...(However,) we have few methods for analyzing data on the sphere or spheroid, and know little about how to model processes on its curved surface ...' (GOODCHILD 1990, p.5f.). The final and perhaps the most important fact why topographic surfaces should be characterized in a formal way is that a formal characterization clearly reveals those concepts which are commonly used in practice but which are seldom or never explicitely stated.

2.TOPOGRAPHIC SURFACES

In almost any geographic or cartographic application functions $f(x,y)$ describing the topography of a given area and associating with each point (x,y) its respective altitude are presumed to be at least twice continuously differentiable. This concept, however, is just an ideal one since, for example, overhanging rocks imply that there is no definite correspondence between certain points and their altitudes or breaklines prevent $f(x,y)$ from being differentiable. In order to apply the powerful tool of calculus, nevertheless, the original concept has to be modified by assuming that the continuously differentiable functions are not the terrain itself but rather sufficiently close approximations of it[2] (cf. WOLF 1988a, 1990).

The question remaining, which seems to be deceptively simple in appearance but which, however, leads rather deeply into abstract mathematics is whether the theoretical requirements of differentiability and continuity of the derivatives suffice for functions to represent realizable topographic surfaces. As will be shown within the next chapters, this must not always be true because such mappings may be endowed with a number of pecularities like degenerate critical points or saddle connections which are extremely unlikely to appear in real-world applications and thus prevent the functions

[1] In order to achieve substantial results in non-topographic applications one will, however, have to ensure that data points are not too scarcely distributed.

[2] This supposition is also valid for mappings describing socio-economic, physical and other phenomena.

from being suitable models for the topography of a given area[3].

Since the detailed investigation of these pecularities requires several concepts from multidimensional calculus, it seems appropriate to repeat some basic definitions and theorems before continuing with the analysis of the addressed phenomena.

Definition 2.1 *A function (mapping)* $f: R^n \rightarrow R$ *is a rule associating with each* $(x_1, x_2, \ldots, x_n) \in R^n$ *a unique element* $f(x_1, x_2, \ldots, x_n) \in R$.

Though the previous definition has been given for the $n-$dimensional case we will restrict ourselves in most instances to two dimensions since this is the commonest case in practical applications. A further advantage of the restriction to functions $f(x, y)$ of only two variables is the fact that these mappings can be easily visualized, thus offering the chance to prefer a geometric approach rather than an abstract one. As a consequence, we will give - whenever possible - not only formal definitions but also geometric interpretations of the concepts being introduced. To start with, let us draw our attention to

Definition 2.2 *The partial derivative* f_x *of a function f(x,y) with respect to the variable x is the derivative of f with respect to x while keepig y constant. The partial derivative* f_y *of f with respect to y is defined in an analogous way. The partial derivatives evaluated at the particular point* (x_0, y_0) *are denoted by* $f_x(x_0, y_0)$ *and* $f_y(x_0, y_0)$ *respectively.*

Geometrically speaking, $f_x(x_0, y_0)$ specifies the tangens of the angle between the tangent to the intersecting curve $f(x, y_0)$ and the line $y = y_0$ parallel to the $x-$axis. To phrase it differently, $f_x(x_0, y_0)$ indicates the slope of the surface $f(x, y)$ at the point (x_0, y_0) in direction to the $x-$axis. It is hardly necessary to point out that $f_y(x_0, y_0)$ can be interpreted in a similar way.

Provided that $f(x, y)$ has partial derivatives at each point $(x, y) \in R^2$, then f_x and f_y are themselves functions of x and y which may also have partial derivatives. These second derivatives (derivatives of order two) are defined recursively by $(f_x)_x = f_{xx}$, $(f_x)_y = f_{xy}$, $(f_y)_x = f_{yx}$ and $(f_y)_y = f_{yy}$. For partial derivatives of order two the following theorem, which is important from a theoretical as well as from a practical point of view, holds[4].

Theorem 2.1 *If the partial derivatives* f_{xy} *and* f_{yx} *of a function f(x,y) are continuous in* R^2 *then* $f_{xy} = f_{yx}$ *in* R^2.

Partial derivatives of order higher than two are defined recursively in an analogous way. We will, however, desist from giving their exact definition since partial derivatives of first and second order are sufficient for the purpose of this paper. Instead we will turn our interest to another point which is of utmost importance for the following chapters and concerns the special arrangement of the partial derivatives of order two in form of a matrix, the so-called Hessian matrix.

[3] For the sake of simplicity we will illustrate these phenomena by examining mappings which are given explicitly and not in form of sparsely distributed data points in combination with an interpolation rule.

[4] For a proof cf. ENDL/LUH (1976, p.185f.).

Definition 2.3 *Let f(x,y) be a function whose partial derivatives* f_{xx}, f_{xy}, f_{yx} *and* f_{yy} *exist. The matrix* $Hf = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}$ *is termed the Hessian matrix of f.*
The Hessian matrix evaluated at a point (x_0, y_0) *is defined by* $\begin{pmatrix} f_{xx}(x_0,y_0) & f_{xy}(x_0,y_0) \\ f_{yx}(x_0,y_0) & f_{yy}(x_0,y_0) \end{pmatrix}$ *and denoted by* $Hf|_{(x_0,y_0)}$.
The determinant det(Hf) of the Hessian matrix Hf is called the Hessian determinant; when evaluated at the point (x_0, y_0) *it is denoted by* $det(Hf)|_{(x_0,y_0)}$.

With the aid of partial derivatives it is now possible to characterize those functions precisely which have been commonly employed for the approximation of topographic surfaces. These mappings are the so-called k–fold continuously differentiable functions whereby in almost any application a value of $k = 2$ has been chosen.

Definition 2.4 *A function f(x,y) is termed k-fold continuously differentiable, or of class* C^k, *if the partial derivatives up to order k exist and are continuous.*
A smooth function is a function of class C^∞.

3.NONDEGENERATE CRITICAL POINTS AND MORSE FUNCTIONS

Critical points[5] representing the peaks, pits and passes of surfaces play a major role not only in cartography but also in a great deal of other scientific applications where they represent either the extrema or the saddles of functions to be maximized or minimized. The importance of the critical points, which are also termed surface-specific points in computer cartography, for this field of research results from the fact that they contain significantly more information than any other point on the surface because they provide information about a specific location as well as about its surrounding (cf. PEUCKER 1973, PFALTZ 1976, PEUCKER/FOWLER/LITTLE/MARK 1978). As a consequence, their employment does not only ease the characterization and visual analysis of the topography of a given area but their application within digital terrain models also results in considerable savings in data capture and data management. Before stating two theorems which allow the classification of the critical points their formal description will be given.

Definition 3.1 *A point* (x_0, y_0) *is a (relative, local) maximum of f(x,y) if and only if* $f(x,y) < f(x_0,y_0)$ *for all* $(x,y) \in U_\epsilon(x_0,y_0)$.
A point (x_0, y_0) *is a (relative, local) minimum of f(x,y) if and only if* $f(x,y) > f(x_0,y_0)$ *for all* $(x,y) \in U_\epsilon(x_0,y_0)$.
A point (x_0, y_0) *is a saddle of f(x,y) if and only if f(x,y) has a local maximum along one line leading through* (x_0, y_0) *and a local minimum along another line leading through* (x_0, y_0).

[5] Unless stated otherwise critical points will be assumed to be nondegenerate.

According to the above definition saddle points are only those points with exactly two ridges (lines connecting passes with peaks) and exactly two courses (lines connecting passes with pits) emanating from them, thus excluding monkey saddles or the like. The following theorem[6] enables the computation as well as the classification of the critical points of a function $f(x,y)$ by applying the concepts of the partial derivatives and the Hessian determinant of $f(x,y)$.

Theorem 3.1 *(x_0, y_0) is a local maximum of a function f(x,y), which is twice continuously differentiable in R^2, if and only if $f_x(x_0,y_0) = f_y(x_0,y_0) = 0$, $det(Hf)|_{(x_0,y_0)} > 0$ and $f_{xx}(x_0,y_0) < 0$ (or equivalently $f_{yy}(x_0,y_0) < 0$).*
(x_0, y_0) is a local minimum of a function f(x,y), which is twice continuously differentiable in R^2, if and only if $f_x(x_0,y_0) = f_y(x_0,y_0) = 0$, $det(Hf)|_{(x_0,y_0)} > 0$ and $f_{xx}(x_0,y_0) > 0$ (or equivalently $f_{yy}(x_0,y_0) > 0$).
(x_0, y_0) is a saddle point of a function f(x,y), which is twice continuously differentiable in R^2, if and only if $f_x(x_0,y_0) = f_y(x_0,y_0) = 0$ and $det(Hf)|_{(x_0,y_0)} < 0$.
(x_0, y_0) is a nondegenerate critical point of a function f(x,y), which is twice continuously differentiable in R^2, if and only if $f_x(x_0,y_0) = f_y(x_0,y_0) = 0$ and $det(Hf)|_{(x_0,y_0)} \neq 0$.

An equivalent characterization of the critical points of a function $f(x,y)$ can be given by examining the eigenvalues of the corresponding Hessian matrix (cf. NACKMAN 1982, p.65 or NACKMAN 1984, p.444f.). The application of eigenvalues has moreover the advantage that they can also be used for the precise mathematical description of ridges, courses, flats, slopes as well as convex and concave hillsides of topographic surfaces (cf. LAFFEY/HARALICK/WATSON 1982, HARALICK/WATSON/LAFFEY 1983). We will, however, refrain from discussing all of these topographic phenomena since this would go far beyond the scope of the present paper.

Theorem 3.2 *Let f(x,y) be twice continuously differentiable in R^2 and $(x_0, y_0) \in R^2$. Further let $f_x(x_0,y_0) = f_y(x_0,y_0) = 0$ and the determinant of the Hessian matrix Hf evaluated at (x_0, y_0) be unequal to zero. Then there is*
a (local) maximum at (x_0, y_0) if the number of negative eigenvalues of $Hf|_{(x_0,y_0)}$ is two,
a saddle at (x_0, y_0) if the number of negative eigenvalues of $Hf|_{(x_0,y_0)}$ is one and
a (local) minimum at (x_0, y_0) if the number of negative eigenvalues of $Hf|_{(x_0,y_0)}$ is zero.

The number of negative eigenvalues of $Hf|_{(x_0,y_0)}$ is also termed the index of (x_0, y_0); thus a maximum is a critical point of index two, a saddle is a critical point of index one, and a minimum is a critical point of index zero. The so-defined index of a critical point may be also interpreted as an 'index of instability (since) a ball displaced slightly from a relative minimum "will roll back" to that minimum. It is a point of stable equilibrium; ... A ball displaced from a saddle point may or may not return to that point of equilibrium, depending on the direction of displacement; while a ball displaced from a

[6] A proof can be found in any standard book on elementary calculus as e.g. in COURANT (1972, p.159f.).

relative maximum is completely unstable' (PFALTZ 1976, p.79, cf. also PFALTZ 1978, p.7f.)[7].

Another advantage of the employment of the eigenvalues of Hf is the fact that this concept may be transfered to n−dimensional differentiable manifolds which represent generalizations of the Euclidean n−space. Formally a manifold is characterized by

Definition 3.2 *An n-dimensional topological manifold is a separable*[8] *metric space in which each point has a neighbourhood homeomorphic*[9] *to* R^n.

An n−dimensional manifold thus represents nothing more than a topological space with the same local properties as the Euclidean n−space. MASSEY (1967, p.1) gives a vivid illustration of the two-dimensional case of this analogy when describing an intelligent bug crawling on a surface (two-dimensional manifold) and being unable to distinguish it from a plane (R^2) due to his limited range of visibility.

The previously defined manifolds, however, must be given some additional structure so that the concept of differentiability has meaning, thus yielding to differentiable manifolds. For the sake of simplicity and because only the concept itself is needed we will drop their formal definition[10] and imagine them as something looking like R^n but being smoothly curved. Examples of two-dimensional differentiable manifolds are the sphere or the torus whereas the cube, the cone or the cylinder are none. With differentiability being specified for mappings defined on manifolds[11] it is possible to investigate not only functions defined on the plane (R^2) but moreover mappings defined on surfaces (two-dimensional manifolds) as e.g. functions describing the distribution of precipitation over the globe because a lot of theoretical results for such mappings can be easily obtained due to the homeomorphic relationships between differentiable manifolds and Euclidean space[12]. Thus the concept of a differentiable manifold as it has been sketched above offers the chance to diminish the deficiency of theoretical knowledge concerning curved surfaces as it has been complained by GOODCHILD (1990, p.5f.) and to counteract his criticism.

Since practice has shown that degenerate critical points are extremely unlikely to occur in real-world applications, functions possessing exclusively

[7] An intuitive classification of the critical points according to their degree of (un)stability has been given by PEUCKER (1973, p.28f.) and WARNTZ/WATERS (1975, p.485f.).

[8] In a topological space, a set $A \subseteq B$ is dense in a set B if $\bar{A} = B$. A topological space C is termed separable if some countable set is dense in C.

[9] Two topological spaces A and B are called homeomorphic if there exists a bijective function $f : A \rightarrow B$ such that both f and f^{-1} are continuous.

[10] The precise mathematical characterization of a differentiable manifold can be found e.g. in GAULD (1982, p.54) or PALIS/de MELO (1982, p.4).

[11] For an adequate definition cf. GAULD (1982, p.60).

[12] For example, it is possible to characterize the critical points by their partial derivatives, to make a distinction between degenerate and nondegenerate ones as well as to classify the latter into maxima, saddles and minima according to the number of negative eigenvalues of the associated Hessian matrix.

nondegenerate critical points have been studied comprehensively by numerous authors[13].

Definition 3.3 *A smooth function is termed a Morse function if all of its critical points are nondegenerate.*

For Morse functions the folllowing four theorems, whose importance will become obvious in the next chapters where the concepts of structural stability and the problem of approximating functions possessing degenerate critical points by mappings without such points will be discussed, hold (cf. ARNOL'D 1972, p.83, PFALTZ 1976, p.83, GAULD 1982, p.118, PALIS/de MELO 1982, p.89, FOMENKO 1987, p.80f.):

Theorem 3.3 *Each Morse function on a compact manifold has only a finite number of critical points; in particular, all of them are distinct.*

Theorem 3.4 *The critical points of a Morse function are always isolated*[14].

Theorem 3.5 *The set of Morse functions is open and dense in the set of all k-fold differentiable functions defined on a manifold.*

Theorem 3.6 *Let f be a Morse function, which is defined on a simply-connected domain bounded by a closed contour line, then the number of minima of f minus the number of saddles of f plus the number of maxima of f equals two.*

The concept of Morse functions - though not explicitly mentioned - has been employed in almost every geographic application since they represent - with one restriction, which will be discussed in Chapter five - the prototype of mappings eligible to characterize topographic surfaces. One exception, however, constitutes the work of PFALTZ (1976, 1978) whose graph theoretic model for the characterization and generalization of topographic surfaces is based explicitly on attributes of Morse functions and thus represents the first attempt to describe those mappings formally which may be regarded as abstract models of the topography of a given area.

4.STRUCTURAL STABILITY

Modern philosophy of science requires that natural science accepts only those theories which can be verified at any time. As a consequence of this metatheoretical view the concept of repeatability saying that the same experiment must give the same result under the same conditions has become fundamental in modern sciences although, strictly speaking, the idea is just an ideal one. Ideal, because it is never possible to guarantee exactly the same conditions by abandoning all external factors even in the most carefully designed experiment. To an even greater extent one is confronted with the

[13] A great deal of the theoretical work is due to Morse (cf. MORSE/CAIRNS 1969).

[14] A critical point is called isolated if sufficiently close to it there exists no other critical point.

problem of repeatability in sciences like geography or cartography. For example, physical-geographic theories are based on measurements taken outside where side-effects are much less controllable than in the physicists' laboratories, or digital terrain models rest on digitized data which are affected by errors due to machine inaccuracy and/or human intervention.

Since the rigorous interpretation of repeatability would make any scientific work impossible the previously described idealized concept has been weakened by tolerating small changes in the conditions under which an experiment is carried out provided that these changes do not affect the result significantly. To phrase it differently, 'what we really expect is not that if we repeat the experiment under precisely the same conditions we will obtain precisely the same results, but rather that if we repeat the experiment under approximately the same conditions we will obtain approximately the same results. This property is known as structural stability ...' (SAUNDERS 1982, p.17). Mathematically, deviations from the ideal experiment which are caused by external factors are represented by perturbation functions and structural stability is the insensitiveness of the mapping or the familiy of mappings describing the experiment to these perturbation functions. The impact of this concept of structural stability for geography and cartography is that in these disciplines questions like the following ones have to be answered: Is a function describing a geographic phenomenon insensitive to small measurement errors and thus structurally stable? Is a family of functions describing a geographic phenomenon over time insensitive to temporal changes and thus structurally stable? Is a mapping representing the underlying continuous model of a digital terrain model insensitive to measurement errors and thus structurally stable?

When using the term 'structural stability', however, one has to distinguish between 'structural stability of a function' (cf. POSTON/STEWART 1978, p.63) and 'structural stability of a family of functions' (cf. POSTON/ STEWART 1978, p.92f., SAUNDERS 1982, p.17f.). In the above-mentioned geographic and cartographic applications of 'structural stability' the first interpretation of the term applies to the first and third examples while the second interpretation applies to the second example. Since in the present paper only the concept of a structurally stable function is of importance we will confine ourselves to this aspect of structural stability and proceed with an example in order to explain it[15].

Let us consider the functions $f_1(x) = x^2$ and $f_2(x) = x^2 + \varepsilon x$ with εx representing a perturbation function. For the derivatives of $f_1(x)$ and $f_2(x)$ holds:

$$\begin{aligned} f_1(x) &= x^2 & \qquad f_2(x) &= x^2 + \varepsilon x \\ f_1'(x) &= 2x & \qquad f_2'(x) &= 2x + \varepsilon \\ f_1''(x) &= 2 & \qquad f_2''(x) &= 2 \end{aligned}$$

[15] For the sake of simplicity we will restrict ourselves thereby to functions of a single variable.

In order to obtain the critical points of the two functions we set the first derivatives to zero, solve the resulting equations with respect to x and examine the second derivates which represent the one-dimensional analogue of the Hessian matrix.

$$\begin{aligned} 2x &= 0 \\ x &= 0 \\ f_1''(0) &= 2 > 0 \end{aligned} \qquad\qquad \begin{aligned} 2x + \varepsilon &= 0 \\ x &= -\frac{\varepsilon}{2} \\ f_2''(-\frac{\varepsilon}{2}) &= 2 > 0 \end{aligned}$$

The above calculations indicate that the perturbation function moves the minimum from $x = 0$ to $x = -\frac{\varepsilon}{2}$ in a way depending smoothly on ε (with ε being an arbitrary small number). The type of the critical point, however, as well as the structure of the graph of $f_1(x)$ in a surrounding of $x = 0$ are not affected by the perturbation (see also Fig. 4.1) and therefore the function is structurally stable at $x = 0$.

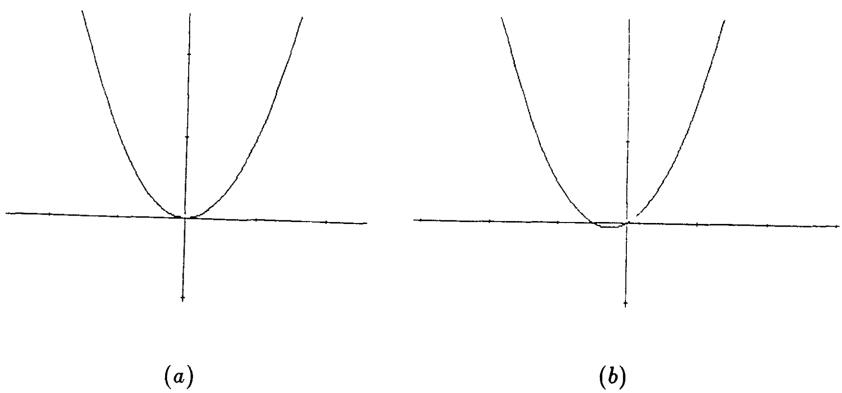


Fig. 4.1 Graphs of the functions (a) $f_1(x) = x^2$ and (b) $f_2(x) = x^2 + \varepsilon x$.

Next let us examine the two mappings $g_1(x) = x^3$ and $g_2(x) = x^3 + \varepsilon x$ with εx representing again a perturbation function. For the derivatives of $g_1(x)$ and $g_2(x)$ holds:

$$\begin{aligned} g_1(x) &= x^3 \\ g_1'(x) &= 3x^2 \\ g_1''(x) &= 6x \end{aligned} \qquad\qquad \begin{aligned} g_2(x) &= x^3 + \varepsilon x \\ g_2'(x) &= 3x^2 + \varepsilon \\ g_2''(x) &= 6x \end{aligned}$$

In order to determine the critical points of $g_1(x)$ and $g_2(x)$ we again set the first derivatives to zero, solve the resulting equations with respect to x

and inspect the second derivatives.

$$\begin{aligned} 3x^2 &= 0 \\ x &= 0 \\ g_1''(0) &= 0 \end{aligned} \qquad\qquad \begin{aligned} 3x^2+\varepsilon &= 0 \\ x_{1,2} &= \pm\sqrt{-\frac{\varepsilon}{3}} \\ g_2''(\pm\sqrt{-\frac{\varepsilon}{3}}) &= \pm 6\sqrt{-\frac{\varepsilon}{3}} \end{aligned}$$

The previous calculations yield to the following interesting result: While the function $g_1(x)$ has a degenerate critical point at $x = 0$, the mapping $g_2(x)$ - which is obtained from $g_1(x)$ by adding the term εx - has no critical points for positive ε but two critical points, namely a local minimum at $x_1 = +\sqrt{-\frac{\varepsilon}{3}}$ and a local maximum at $x_2 = -\sqrt{-\frac{\varepsilon}{3}}$, for negative ε thus showing an irregular unstable behaviour. Illustrations of the function $g_2(x) = x^3 + \varepsilon x$ for different values of ε are depicted in Fig. 4.2.

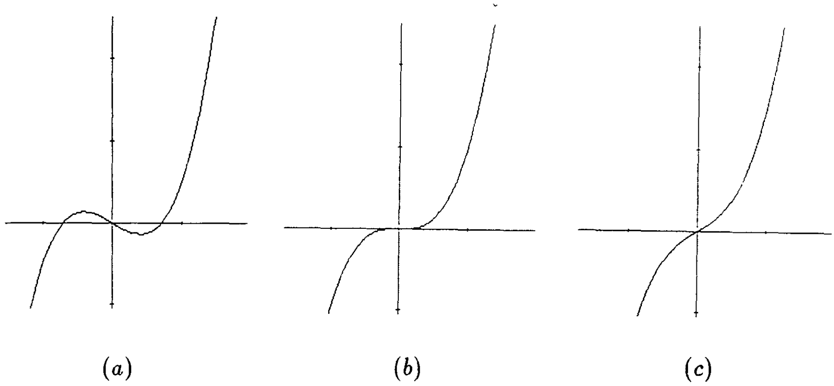


Fig. 4.2 Graphs of the function $g_2(x) = x^3 + \varepsilon x$ for (a) $\varepsilon < 0$, (b) $\varepsilon = 0$ and (c) $\varepsilon > 0$.

The different behaviour of the two functions $f_1(x)$ and $g_1(x)$ in a surrounding of $x = 0$ can be explained by the following theorem (cf. POSTON/STEWART 1978, p.63f.).

Theorem 4.1 *A critical point is structurally stable if and only if it is nondegenerate.*

In the above example $f_1(x)$ has a nondegenerate critical point at $x = 0$ while $g_1(x)$ has a degenerate one at this location. In the first case, as a consequence of the structural stability induced by the nondegenerate critical point the perturbation function does not change the type of the point but only moves its location. In the second case, however, the degeneracy of the critical point causes structural instability resulting in a change of the type of the critical point.

A direct consequence of the last theorem is the following one which shows once more the importance of Morse functions (cf. POSTON/STEWART 1978, p.70f.).

Theorem 4.2 *Morse functions are structurally stable.*

At the beginning of this chapter the importance of structural stability for scientific work in general has been indicated. At this point its importance for geography and cartography will be demonstrated in the light of some possible applications. The most essential one will certainly concern those fields in geographic and cartographic research where the systems approach is already well established as e.g. in ecology, climatology, demography etc. and functions or families of functions are used to describe ecosystems, weather, population dynamics etc. The major question to be answered in the above examples is whether a given system is stable or unstable over time and in the latter case if it will explode or collapse. A second field of applications comprises the analysis of functions describing cartographic and geographic phenomena like terrain, population density, accessibility, temperature and the like. The question to be answered in this context is which data points are best selected so that the functions obtained are structurally stable. It can be assumed, however, that those mappings that are derived from surface-specific points which are taken to be nondegenerate will produce results being superior to all others. Finally, a third point worth mentioning in this connection is the analysis of structural stability due to measurement errors caused by machine inaccuracy and/or human intervention - a problem which will have to be tackled in combination with the aid of statistics.

5.DEGENERATE CRITICAL POINTS AND SADDLE CONNECTIONS

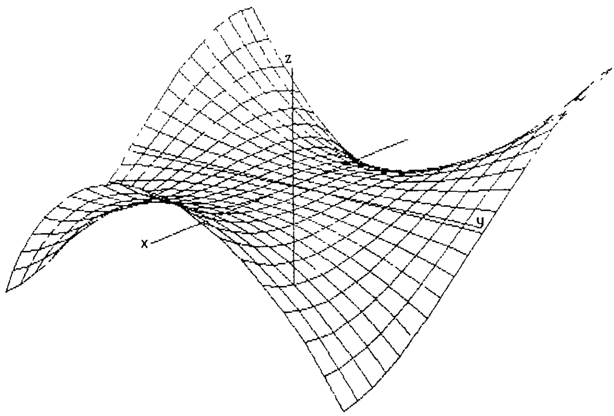
Degenerate critical points form - besides saddle connections - part of those phenomena which prevent continuously differentiable mappings from being suitable models for the topography of a given area. The reason is that degenerate critical points are - according to Theorem 4.1 - structurally unstable and thus unlikely to appear in real-world applications since any perturbation would immediately destroy them.

Formally, a degenerate critical point (x_0, y_0) is characterized by the fact, that the partial derivatives $f_x(x_0, y_0)$ and $f_y(x_0, y_0)$ as well as the Hessian determinant $det(Hf)|_{(x_0,y_0)}$ are zero. Some examples of functions possessing degenerate critical points are depicted in Fig. 5.1.

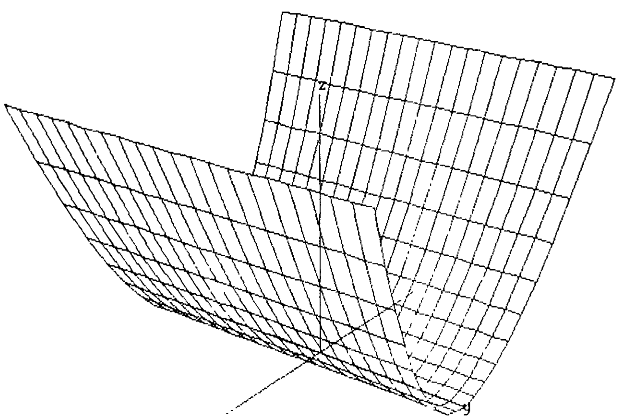
Though its definition sounds deceptively simple, degeneracy is a multifaceted phenomenon with different levels to be distinguished. A first subdivision can be made into isolated degenerate critical points and non-isolated ones with the latter being extremely uncommon[16] (cf. POSTON/STEWART

[16] For this reason they are excluded from further consideration.

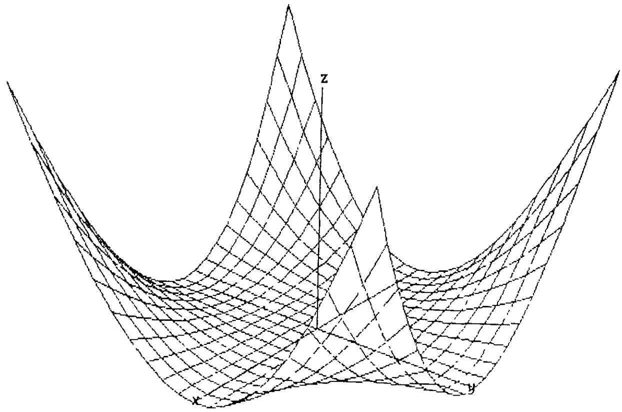
1978, p.53.). From the above sketched functions $f(x,y)$ has an isolated degenerate critical point at (0,0) while $g(x,y)$ has non-isolated ones along the x-axis and $h(x,y)$ along the x- and the y-axes. Besides this subdivision of the critical points another one can be made according to their degree of degeneracy (cf. FOMENKO 1987, p.80) which is explained next.



(a)



(b)



(c)

Fig. 5.1 Some functions having degenerate critical points: (a) $f(x,y) = x^3 - 3xy^2$ (monkey saddle at (0,0)); (b) $g(x,y) = x^2$ (pig-trough with degenerate critical points occuring along the x–axis); (c) $h(x,y) = x^2y^2$ (crossed pig-trough with degenerate critical points occuring along the x– and the y–axes).

Definition 5.1 *The degree of degeneracy of a critical point* (x_0, y_0) *is equivalent to the number of zero eigenvalues of* $Hf|_{(x_0,y_0)}$.

To illustrate this concept let us examine the two functions $f(x,y) = x^3 - 3xy^2$ and $g(x,y) = \frac{x^3}{3} - \frac{y^2}{2}$. By setting the first partial derivatives to zero and inspecting the second partial derivatives it can be shown that both $f(x,y)$ and $g(x,y)$ possess a degenerate critical point at (0,0).

$$\begin{aligned} f(x,y) &= x^3 - 3xy^2 \\ f_x(x,y) &= 3x^2 - 3y^2 \\ f_y(x,y) &= -6xy \\ f_{xx}(x,y) &= 6x \\ f_{xy}(x,y) = f_{yx}(x,y) &= -6y \\ f_{yy}(x,y) &= -6x \end{aligned}$$

$$\begin{aligned} g(x,y) &= \frac{x^3}{3} - \frac{y^2}{2} \\ g_x(x,y) &= x^2 \\ g_y(x,y) &= -y \\ g_{xx}(x,y) &= 2x \\ g_{xy}(x,y) = g_{yx}(x,y) &= 0 \\ g_{yy}(x,y) &= -1 \end{aligned}$$

The Hessian matrices of the two mappings run therefore

$$Hf = \begin{pmatrix} 6x & -6y \\ -6y & -6x \end{pmatrix} \qquad\qquad Hg = \begin{pmatrix} 2x & 0 \\ 0 & -1 \end{pmatrix}$$

and when evaluated at the critical point (0,0)

$$Hf|_{(0,0)} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \qquad\qquad Hg|_{(0,0)} = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}$$

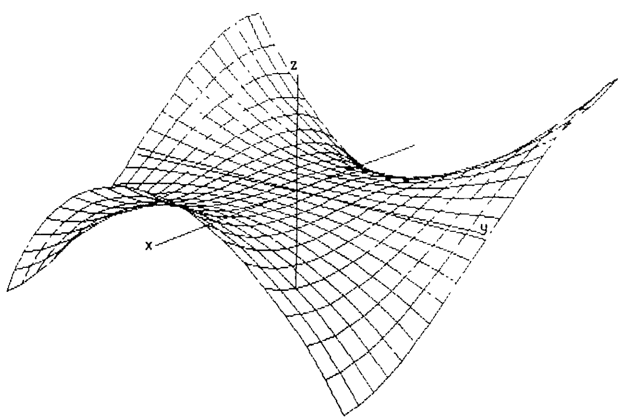
In order to obtain the eigenvalues of the two matrices we solve the corresponding characteristic polynoms

$$(0-\lambda)(0-\lambda) \;=\; 0 \qquad\qquad (0-\lambda)(-1-\lambda) \;=\; 0$$

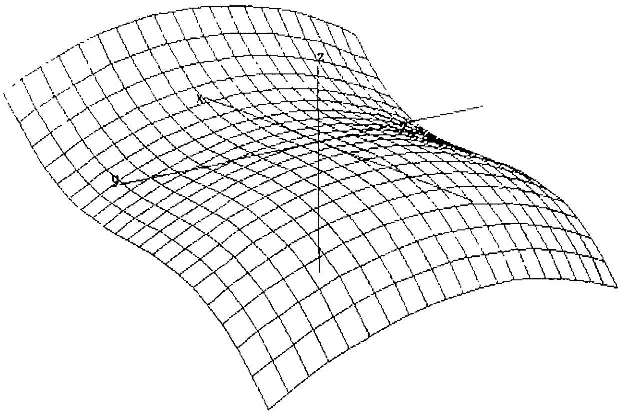
yielding

$$\lambda_{1,2} \;=\; 0 \qquad\qquad \begin{aligned} \lambda_1 &= 0 \\ \lambda_2 &= -1 \end{aligned}$$

Thus, in the first case the number of zero eigenvalues and therefore the degree of degeneracy of the critical point (0,0) is two, whereas in the second case the degree of degeneracy of (0,0) is one. Illustrations of the two functions in a surrounding of this location can be found in Fig. 5.2.



(a)



(b)

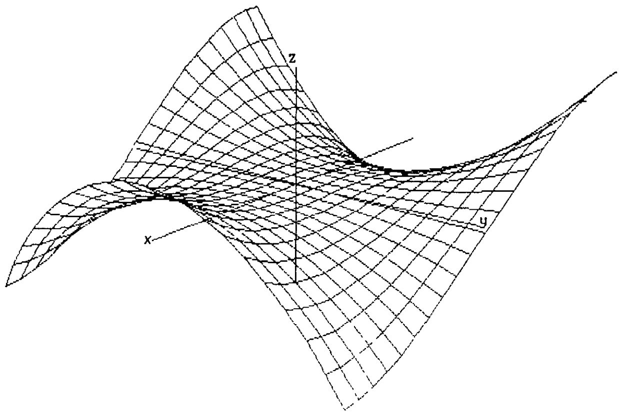
Fig. 5.2 Graphs of the functions (a) $f(x,y) = x^3 - 3xy^2$ and (b) $g(x,y) = \frac{x^3}{3} - \frac{y^2}{2}$.

Another point of interest concerning degeneracy is the question whether functions possessing degenerate critical points can be approximated accurately enough by mappings without such points, or in other words if it is possible to substitute degenerate critical points by nondegenerate ones. In order to answer this question let us recall that degenerate critical points are structurally unstable and that the set of Morse functions is open and dense in the set of all differentiable mappings defined on a manifold. It can be proved that due to these two properties the question asked earlier can be answered affirmatively since the following theorem[17] holds.

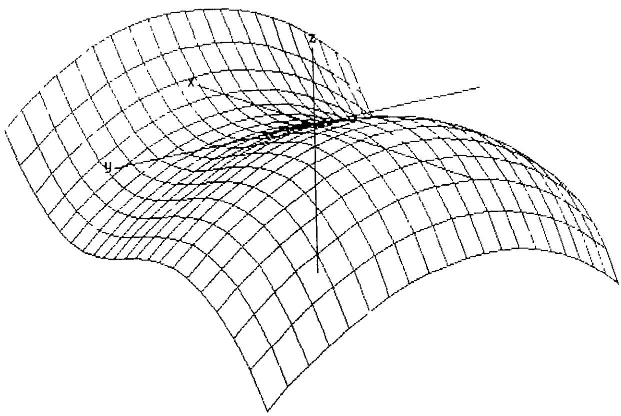
Theorem 5.1 *If a function has a degenerate critical point, then by an arbitrarily small shift of the function it can be ensured that the complicated singularity is dispersed into several nondegenerate ones.*

The above theorem, however, does not provide any information about the number nor about the types of the nondegenerate critical points one obtains when splitting a degenerate one. The following examples illustrate two possible cases that might occur when mappings are interfered by perturbation functions by means of the two mappings $f(x,y) = x^3 - 3xy^2$ and $g(x,y) = \frac{x^3}{3} - \frac{y^2}{2}$ both possessing a degenerate critical point at (0,0) (see Fig. 5.2). Deformations of $f(x,y)$ and $g(x,y)$ by the perturbation functions εy and εx respectively yield $\tilde{f}(x,y) = x^3 - 3xy^2 - \varepsilon y$ and $\tilde{g}(x,y) = \frac{x^3}{3} - \frac{y^2}{2} - \varepsilon x$. When

[17] An exact proof of this theorem which is rather complicated and requires several concepts like transversality, jet-spaces etc. from such branches of abstract mathematics as differential topology or catastrophy theory can be found in ARNOL'D (1972, p.65).



(a)



(b)

Fig. 5.3 Graphs of the functions (a) $\tilde{f}(x, y) = x^3 - 3xy^2 - \varepsilon y$ and (b) $g(x, y) = \frac{x^3}{3} - \frac{y^2}{2} - \varepsilon x$.

setting the first partial derivatives of $\tilde{f}$ and $\tilde{g}$ to zero, solving the resulting systems of equations with respect to x and y, and inspecting the second partial derivatives it becomes apparent that the degenerate critical point (0,0) of $f(x,y)$ is substituted by two nondegenerate saddles with locations at $(\sqrt{\frac{\varepsilon}{6}}, -\sqrt{\frac{\varepsilon}{6}})$ and $(-\sqrt{\frac{\varepsilon}{6}}, \sqrt{\frac{\varepsilon}{6}})$ respectively, whereas the degenerate critical point (0,0) of $g(x,y)$ is substituted by a nondegenerate saddle at $(\sqrt{\varepsilon}, 0)$ and a local maximum at $(-\sqrt{\varepsilon}, 0)$. The visualization of the effect of approximating a function having a degenerate critical point by a mapping without such points can be achieved by comparing Fig. 5.2 and Fig. 5.3 with the latter depicting the graphs of $\tilde{f}(x,y)$ and $\tilde{g}(x,y)$.

It can easily be demonstrated that degenerate critical points are not the only phenomena inducing structural instability but saddle connections will cause it, too. However, it has been proven that saddle connections may always be broken up by perturbation functions which have to be chosen in a convenient way (cf. GUCKENHEIMER/HOLMES 1983, p.60ff.). As a consequence of this result and Theorem 5.1 it can be concluded that Morse functions without saddle connections are the most suitable mappings to describe topographic surfaces because, on the one hand, they possess only structural stable elements like nondegenerate critical points but are, on the other hand, also eligible to approximate accurately enough structural unstable elements like degenerate critcal points and saddle connections.

6.CONCLUSION

In the present paper the characterization of those mappings which may be regarded as abstract models of topographic surfaces has been attempted. The importance of a characterization like this is derived from the fact that differentiability and continuity of the derivatives do not suffice for functions to represent realizable topographic surfaces because continuously differentiable mappings may nevertheless be endowed with pecularities which are unlikely to appear in reality. An analysis of these pecularities, however, reveals that they are primarily due to structural instability of the respective functions - a phenomenon induced by degenerate critical points or saddle connections. Therefore it has been concluded that mappings describing the topography of a given area should be Morse functions without saddle connections. It should be emphasized, however, that the results obtained in this article represent only the first step in the formal characterization of the topography of a given area because a great deal of important phenomena like junctions of channels and ridges have not been considered. The analysis of these phenomena and its incorporation into a general framework of spatial data management will have to be the subject of future research.

7.REFERENCES

ARNOL'D, V. I., 1972. Lectures on Bifurcations in Versal Families. In: Russian Mathematical Surveys, p.54 - 123.

BOUDRIAULT, G., 1987. Topology of the TIGER File. In: Proceedings of AUTOCARTO 8, p.258 - 263.

COURANT, R., 1972. Vorlesungen ueber Differential- und Integralrechnung 2. Berlin, Heidelberg, New York.

ENDL, K./LUH, W., 1976. Analysis II. Wiesbaden.

FOMENKO, A. T., 1987. Differential Geometry and Topology. New York, London.

GAULD, D. B., 1982. Differential Topology. An Introduction. New York, Basel.

GOODCHILD, M. F., 1990. Spatial Information Science. In: Proceedings of the Fourth International Symposium on Spatial Data Handling, p.3 - 12.

GUCKENHEIMER, J./HOLMES, P., 1983. Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields. New York, Berlin, Heidelberg, Tokyo.

HARALICK, R. M./WATSON, L. T./LAFFEY, T. J., 1983. The Topographic Primal Sketch. In: The International Journal of Robotics Research, p.50 - 72.

LAFFEY, T. J./HARALICK, R. M./WATSON, L. T., 1982. Topographic Classification of Digital Image Intensity Surfaces. In: Proceedings of the IEEE Workshop on Computer Vision, Theory and Control, p.171 - 177.

MASSEY, W. S., 1967. Algebraic Topology: An Introduction. New York, Chicago, San Francisco, Atlanta.

MORSE, M./CAIRNS, S. S., 1969. Critical Point Theory in Global Analysis and Differential Topology. New York, London.

NACKMAN, L. R., 1982. Three-dimensional Shape Description Using the Symmetric Axis Transform. Ph.D. dissertation. Department of Computer Science, University of North Carolina, Chapel Hill.

NACKMAN, L. R., 1984. Two-Dimensional Critical Point Configuration Graphs. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, p.442 - 450.

PALIS, Jr. J./de MELO W., 1982. Geometric Theory of Dynamical Systems. New York, Heidelberg, Berlin.

PEUCKER, T. K., 1973. Geographic Data Structures. Progress Report After Year One. Technical Report #1, 'Geographic Data Structures' Project, ONR Contract N00014-73-C-0109.

PEUCKER, T. K./CHRISMAN, N., 1975. Cartographic Data Structures. In: The American Cartographer, p.55 - 69.

PEUCKER, T. K./FOWLER, R. J./LITTLE, J. J./MARK, D. M., 1978. The Triangulated Irregular Network. In: Proceedings of the American Society of Photogrammetry. Symposium on Digital Terrain Models (DTM), St. Louis, p.516 - 540.

PEUQUET, D., 1983. A Hybrid Structure for the Storage and Manipulation of Very Large Spatial Data Sets. In: Computer Vision, Graphics, and Image Processing, p.14 - 27.

PFALTZ, J. L., 1976. Surface Networks. In: Geographical Analysis, p.77 - 93.

PFALTZ, J. L., 1978. Surface Networks, an Analytic Tool for the Study of Functional Surfaces. Final report on NSF Grant OCR-74-13353.

POSTON, T./STEWART, I., 1978. Catastrophy Theory and Its Applications. London, San Francisco, Melbourne.

SALGÉ, F./SCLAFER, M. N., 1989. A Geographic Data Model Based on HBDS Concepts: The IGN Cartographic Data Base Model. In: Proceedings of AUTOCARTO 9, p.110 - 117.

SAUNDERS, P. T., 1982. An Introduction to Catastrophe Theory. Cambridge.

WARNTZ, W./WATERS, N., 1975. Network Representations of Critical Elements of Pressure Surfaces. In: Geographical Review, p.476 - 492.

WEIBEL, R., 1989. Konzepte und Experimente zur Automatisierung der Reliefgestaltung. Dissertation. Philosophische Fakultaet, Universitaet Zuerich.

WOLF, G. W., 1988a. Generalisierung topographischer Karten mittels Oberflaechengraphen. Dissertation. Institut fuer Geographie, Universitaet Klagenfurt.

WOLF, G. W., 1988b. Weighted Surface Networks and their Application to Cartographic Generalization. In: BARTH, W. (ed.), 1988. Visualisierungstechniken und Algorithmen. Berlin, Heidelberg, p.199 - 212.

WOLF, G. W., 1989. A Practical Example of Cartographic Generalization Using Weighted Surface Networks. In: DOLLINGER, F./STROBL, J. (eds.), 1989. Angewandte Geographische Informationstechnologie (= Salzburger Geographische Materialien, Heft 13), p.125 - 143.

WOLF, G. W., 1990. Metric Surface Networks. In: Proceedings of the Fourth International Symposium on Spatial Data Handling, p.844 - 856.