THE TEACHING OF MATHEMATICS

2011, Vol. XIV, 2, pp. 137–145

# THREE MANIFESTATIONS OF MORSE THEORY IN TWO DIMENSIONS

**Vladimir N. Grujić**

*Dedicated to Professor Milosav Marjanović on the occasion of his 80th birthday*

**Abstract.** In this article we present main notions and ideas of Morse theory in two dimensions, adjusted to school teachers and their talented students. We count numbers of critical points of different types and obtain interesting results about plane curves, mountainous landscapes and planets. We also derive the Euler formula for polyhedra.

*ZDM Subject Classification*: I25; *AMS Subject Classification*: 97I20.

*Key words and phrases*: Critical point; Morse function; polyhedron.

## 1. Introduction

Topology is a modern branch of mathematics that studies properties of spaces invariant under continuous deformations. In contrast to geometry that is concerned with metric properties of spaces, in topology the metric is irrelevant. In that sense topological properties are more essential and describe more deeply qualitative characteristics of spaces. For instance, from geometrical point of view a circle and a square are quite different figures but they share some common properties as dimension and both divide the plane into two parts. They can be transformed into each other by continuous deformation. Transformations of this kind are called *homeomorphisms*. They are the main topological equivalences of spaces and from topological point of view homeomorphic spaces are the same. Topological properties have wide applications in all areas of modern mathematics, as well as in physics, computer science and even in economy. This makes topology the unique discipline that reflects a unifying principle of modern mathematics. We present here one of the most beautiful topological theories that influenced development of topology and that is ever young and actual. This is Morse theory[1]. It is concerned with shapes of spaces called manifolds and with critical points of real functions on them. The unsurpassed exposition of the subject is the famous Milnor's book [1], one of the most cited books in mathematics. We start with three geometric problems referred to as "pictures".

PICTURE 1. Count the number of minima $m$ and maxima $M$ of a closed plane curve $\gamma$ in the general position without selfintersections. Are they always the same, namely is it $\chi = M - m = 0$?

[1]Marston Morse, 1892–1977, American mathematician, stated his theory in mid 30's of the last century.

Picture 2. Given a mountainous landscape $\mathcal{M}$, count the number $m$ of points of minimal height (basins), the number $M$ of points of maximal height (peaks) and the number $p$ of saddle points (passes). Calculate $\chi = m - p + M$. Is it always $\chi = 1$ for every landscape?

Picture 3. Let $\Pi$ be a convex polyhedron with $T$ vertices, $I$ edges and $S$ faces. The famous Euler formula says that $\chi = T - I + S = 2$ for any convex polyhedron $\Pi$. Check this for Platonic solids.

We see that all three pictures say something about the number $\chi$ that does not depend on the metric shape of involved spaces. Hence they are topological invariants in nature. We prove this in what follows, but let us first introduce the main actors of Morse theory.

## 2. Critical points

We consider real functions of $n$ variables $f : \mathbb{R}^n \to \mathbb{R}$ and for the beginning we restrict ourselves to the cases when $n = 1, 2$. The nicest class of such functions, called analytic functions, are those that can be expanded into power series of the form

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots \tag{1}$$

$$f(x, y) = a_0 + (a_{11}x + a_{12}y) + (a_{21}x^2 + 2a_{22}xy + a_{23}y^2) + a_{31}x^3 + \cdots$$

They are generalization of polynomials, the simplest and the well-known class of real functions.

Let $f : \mathbb{R} \to \mathbb{R}$ be an analytic function of the form (1). We say that $f$ has a critical point at the origin $x = 0$ if $a_1 = 0$ in its series expansion. That critical point is called degenerate if $a_2 = 0$ and nondegenerate if $a_2 \neq 0$. These conditions are equivalent to $f'(0) = f''(0) = 0$ for degenerate and $f'(0) = 0, f''(0) \neq 0$ for nondegenerate critical point. Value of the function in the critical point $a_0 = f(0)$ we call the critical value. The type of a critical point does not depend on the critical value, so we may suppose that $a_0 = 0$.

The main contribution to the behavior of function $f$ near the nondegenerate critical point comes from the second summand $a_2 x^2$. The type of the nondegenerate critical point depends on the sign of the coefficient $a_2$ in the way that if $a_2 > 0$ it is the local minimum and if $a_2 < 0$ it is the local maximum of function $f$ near the origin.

We call a critical point *stable* if it cannot be destructed or new critical points cannot be created by a small deformation of function $f$ in its small neighborhood. The degenerate critical point is not stable. For example, let $a_0 = a_1 = a_2 = 0$ and $a_3 > 0$ in the series expansion of function $f$. Then function $f$ near the origin looks like the cubic line $y = a_3 x^3$. It can be slightly deformed by the small parameter $t \neq 0$ to obtain $y = a_3 x^3 + tx$, which does not have critical points for $t > 0$ and has two critical points for $t < 0$ near the origin. This deformation is called the *creation* or the *destruction* of the critical point. Unlike this situation, the nondegenerate critical point is stable. That is the reason why we are interested only in nondegenerate critical points.

Definition. A function $f$ with only nondegenerate critical points is called a *Morse function*.

Now we can explain what we have meant by "the general position of the closed curve $\gamma$" in Picture 1. The curve $\gamma$ can be described by the equation $\varphi(x, y) = 0$. If the function $\varphi$ is analytic we say that $\gamma$ is a smooth curve. Define the function $f : \gamma \to \mathbb{R}$ on the smooth curve $\gamma$ by $f(x, y) = y$. Geometrically it measures the height of point $(x, y) \in \gamma$ according to $y$-axis. We call $f$ the height function. We say that the curve $\gamma$ is in a general position if its height function is a Morse function. The proof of equality of the number of local minima $m$ and local maxima $M$ follows from the simple fact that any minimum has to be followed by a maximum and vice versa.

We will now treat a reacher case of the dimension $n = 2$. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be an analytic function of the form (2). Similarly as in the case of functions of one variable we define the notion of (non)degenerate critical points and critical values for functions of two variables. The origin $(x, y) = (0, 0)$ is a critical point of function $f$ if $a_{11} = a_{12} = 0$ in the series expansion (2). For a reader who is familiar with partial derivatives, this is equivalent to the condition

$$\frac{\partial f}{\partial x}(0,0) = \frac{\partial f}{\partial y}(0,0) = 0.$$

The value $a_0 = f(0, 0)$ of function $f$ at the critical point is a critical value and we may suppose that $a_0 = 0$. We call noncritical values regular. The behavior of function $f$ near the critical point depends on the second term

$$q(x, y) = a_{21}x^2 + 2a_{22}xy + a_{23}y^2$$

of its series expansion. This term is called the quadratic form (see [4] to learn more about quadratic functions) associated to function $f$ at the critical point. It can be written in the form of matrix multiplication

$$q(x, y) = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a_{21} & a_{22} \\ a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

The $2\times 2$-matrix $H = \begin{pmatrix} a_{21} & a_{22} \\ a_{22} & a_{23} \end{pmatrix}$ is called the Hessian of function $f$ in the critical point. We recognize coefficients of the Hessian as the second partial derivatives of function $f$,

$$a_{21} = \frac{\partial^2 f}{\partial x^2}(0,0), \quad a_{22} = \frac{\partial^2 f}{\partial x \partial y}(0,0), \quad a_{23} = \frac{\partial^2 f}{\partial y^2}(0,0).$$

Let $\Delta = detH = a_{21}a_{23} - a_{22}^2$ be the determinant of Hessian matrix. The quadratic form $q(x, y)$ is called degenerate if $\Delta = 0$ and nondegenerate if $\Delta \neq 0$. We distinct the character of the critical point by this rule. We can transform coordinates in a small neighborhood of the nondegenerate critical point to get a quadratic form of the Morse function $f$. This guaranties the following

Morse lemma. *There is a transformation of coordinates* $(x,y)\mapsto(X,Y)$ *in a small neighborhood of a nondegenerate critical point after which the Morse function* $f$ *has a simple form*

$$f(X,Y)=\pm X^2\pm Y^2,$$

*where signs depend on the Hessian matrix* $H$.

As a consequence we see that the nondegenerate critical point is isolated. It is the only critical point in some of its neighborhoods. The number of minuses $\lambda$ in the above presentation of function $f$ in new coordinates $(X,Y)$ is called the *index* of the critical point. The dependence of signs on the Hessian matrix geometrically corresponds to the intuitive notion that the index $\lambda$ is the number of independent directions in which function $f$ decreases. Consequently we have three types of nondegenerate critical points of Morse functions of two variables: minimum for $\lambda=0$, saddle point for $\lambda=1$ and maximum for $\lambda=2$. They are depicted in Figure 1.

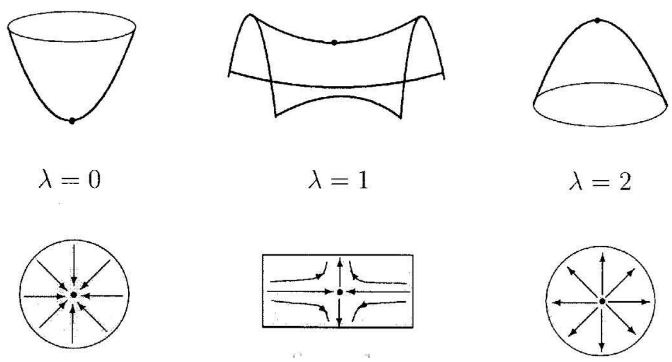



Fig. 1

We introduce the notion of gradient flow lines of the Morse function $f$. At any point of the domain we put a vector in the direction of the fastest growth of function $f$ with the magnitude equal to the slope of function $f$ in that direction. Such assignments of vectors to points of the domain we call the *gradient field* of function $f$ and denote by $\operatorname{grad} f$. Stationary points of the gradient field, i.e. points where it vanishes are exactly critical points of function $f$. Trajectories of the gradient field are curves in the domain for which the tangent vector at each point is equal to the gradient vector at that point. We call these trajectories *gradient flow lines.* Through each nonstationary point the unique gradient flow line passes. Gradient flow lines have another important characteristic. Define level sets of function $f$ as curves $f(x,y)=c$ for different values of $c$. Gradient flow lines are perpendicular to level sets. We usually draw the opposite field $-\operatorname{grad} f$ and reverse directions of flow lines. In Figure 1 we depicted gradient flow lines around critical points of different indexes.

### 3. On hills and lakes[2]

We return now to Picture 2. The mountainous landscape $\mathcal{M}$ can be described by an equation $z = f(x, y)$ in the 3-space $\mathbb{R}^3$,

$$\mathcal{M} = \{\, (x, y, z) \in \mathbb{R}^3 \mid z = f(x, y),\ (x, y) \in D \,\},$$

where $D$ is some simple closed bounded plane domain, such that on its boundary $\partial D$ function $f$ takes the zero value $f(x, y) = 0$, $(x, y) \in \partial D$. We can imagine the mountainous landscape $\mathcal{M}$ as an island in the plane ocean. In that vision the boundary $\partial D$ will be the coast of the island. We are looking at the height function $h : \mathcal{M} \to \mathbb{R}$ defined by $h(x, y, z) = z$, $(x, y, z) \in \mathcal{M}$ which measures elevations of island's points with respect to the $z$-axis. We say that $\mathcal{M}$ is in a general position if all critical points of the height function are nondegenerate and none of the critical points belongs to the coast of the island. It means that we have only peaks, passes and basins inside the island.

For reasons of simplicity of the proof we may suppose that all basins have the same depth, all peaks have the same height and that all passes are on different levels. This can be achieved by changing the function $z = f(x, y)$ locally around critical points with no effects on their types and total number. These local changes of the landscape $\mathcal{M}$ are indicated in Figure 2.

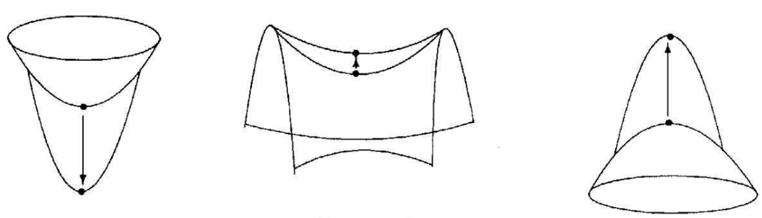


Fig. 2

Then the height function $h$ has the common minimal value $\mu$ at points of minima, critical values $\alpha_1 < \cdots < \alpha_p$ at saddle points and the common maximal value $\nu$ at points of maxima. Functions with these properties we call *proper Morse functions.* Also suppose that the common depth of basins are above the ocean level, i.e. $\mu > 0$.

Define a level set $\mathcal{M}^t = h^{-1}(\{t\})$ as the set of all points with elevations $t$. For a regular value $t$, the connected components of the level set $\mathcal{M}^t$ are closed curves

[2]This is a paraphrase of the title of James Clark Maxwell's article *On hills and dales* [2]

without selfintersections[3]. For a critical value $t$, the connected components of level set $\mathcal{M}^t$ are either isolated points (corresponding to maxima and minima) or closed curves with double points (corresponding to saddle points). The sublevel region $\mathcal{M}^{\leq t}$ is defined as the set of all points of the island $\mathcal{M}$ with the height at most $t$,

$$\mathcal{M}^{\leq t} = \{(x, y, z) \in \mathcal{M} | f(x, y) \leq t\}.$$

Similarly, the level sector $\mathcal{M}^{[a,b]}$ is the set of all points with the height between the given levels,

$$\mathcal{M}^{[a,b]} = \{(x, y, z) \in \mathcal{M} | a \leq f(x, y) \leq b\}.$$

We want to understand how these regions change when the parameter $t$ increases. The main tool needed for this analysis is provided by the following theorem.

THEOREM 1. *Suppose that $a < b$ are real parameters such that there are no critical values between $a$ and $b$. Then the regions $\mathcal{M}^{\leq a}$ and $\mathcal{M}^{\leq b}$ have the same shape, topologically speaking they are homeomorphic. Also, the level sector $\mathcal{M}^{[a,b]}$ is homeomorphic to the product $\mathcal{M}^a \times [0, 1]$ which we call a band along the level set $\mathcal{M}^a$.*

This means that the main changes of shapes of sublevel regions $\mathcal{M}^{\leq t}$ may occur only when the parameter $t$ passes the critical values. For this reason we choose a discrete set of values for the parameter

$$t : \quad \mu < t_1 < \alpha_1 < t_2 < \alpha_2 < \cdots < \alpha_{p-1} < t_p < \alpha_p < t_{p+1} < \nu$$

and look what is the effect on the sublevel regions

$$\mathcal{M}^{\leq t_1} \subset \mathcal{M}^{\leq t_2} \subset \cdots \subset \mathcal{M}^{\leq t_{p+1}}.$$

It is obvious that $\mathcal{M}^{\leq t} = \emptyset$ for $t < 0$ and $\mathcal{M}^{\leq t} = \mathcal{M}$ for $t > \nu$. $\mathcal{M}^0$ is the island's coast line $\partial D$. From Theorem 1 we have that the level sector $\mathcal{M}^{[0,t]}$, for any $t < \mu$, is homeomorphic to the band along the coast line $\mathcal{M}^0 \times [0, 1]$. We may think about this band as the island's beach. What happens when $t$ passes the critical value $\mu$? The region $\mathcal{M}^{\leq t_1}$ contains the island's beach and $m$ separate basins. We imagine that these basins are filled with water and form $m$ lakes on the island $\mathcal{M}$.

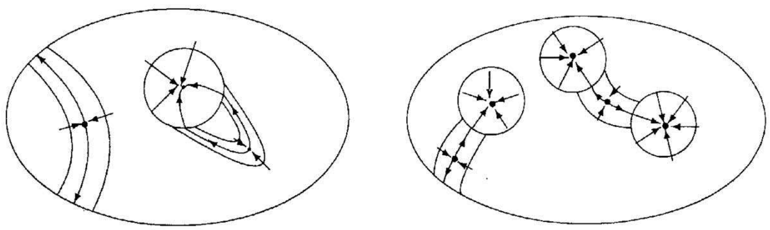
From now on we think about connected components of regions $\mathcal{M}^{\leq t}$ that are *separated* from the ocean as lakes and about connected components of sectors $\mathcal{M}^{[t,\nu]}$ as hills. We are interested in numbers of lakes and hills when the parameter $t$ increases. By Theorem 1 these numbers could be changed only when $t$ passes through critical values. We take the case $t = t_1$ as the start position: $m$ lakes and one hill.

How numbers of lakes and hills change after passing the critical value $\alpha_1$ at the first saddle point? The answer is given by the following theorem.

THEOREM 2. *Suppose that $a < b$ are real parameters such that there is only one critical value between $a$ and $b$ corresponding to the saddle point. Then the region $\mathcal{M}^{\leq b}$ is obtained from the region $\mathcal{M}^{\leq a}$ by attaching a band along its opposite sides.*

[3]known as isohypses on geographic charts

According to Theorem 2 the region $\mathcal{M}^{\leq t_2}$ is obtained from the region $\mathcal{M}^{\leq t_1}$ by attaching a band along its opposite sides. We distinguish four ways to attach a band according to where its sides are attached. We collect all these ways into two cases. The first case is when both sides are attached either on a single lake or on the ocean. The second case is when they are attached on different lakes or one side is attached to the ocean and the other one to some lake.



the first case the second case

Fig. 3

In the first case the number of lakes remains the same, but the number of hills increases by one. In the second case the number of lakes decreases by one and the number of hills remains the same. The same is true for any critical value $\alpha_j$ and the corresponding regions $\mathcal{M}^{\leq t_j}$ and $\mathcal{M}^{\leq t_{j+1}}$. Finally, we approach the last case $t = t_{p+1}$ as the final position: no lakes and $M$ hills corresponding to $M$ peaks. We observe that the number of lakes decreases by $m$ and the number of hills increases by $M - 1$ moving from the start position $t = t_1$ to the final position $t = t_{p+1}$. We conclude that there must be $m + M - 1$ attaching bands corresponding to saddle points. Thus, we have proved the equality stated in Picture 2,

$$m - p + M = 1. \tag{3}$$

## 4. Polyhedra

Now we address Picture 3. Let us consider the 2-dimensional sphere $S^2$ and Morse functions on it. We may imagine the surface of some planet $\mathcal{P}$ and a distance function $d : \mathcal{P} \to \mathbb{R}$ from some of the interior points $O$ of the planet $\mathcal{P}$. Suppose that $d$ is a Morse function on the surface of $\mathcal{P}$. This means that the planet $\mathcal{P}$ has only basins, passes and peaks relative to function $d$. Let us choose the deepest basins, the point on the surface of $\mathcal{P}$ where function $d$ is approaching the absolute minimal value, and surround it in a small neighborhood by water. We may think about a water neighborhood of the deepest basin as an ocean on the planet $\mathcal{P}$, small one, but still an ocean. The rest of the planet $\mathcal{P}$ we will treat as the island. Let $m', p$ and $M$ be numbers of minima, saddles and maxima of function $d$ on the island. According to formula (3) we have that $m' - p + M = 1$. As we have only one more critical point, namely the deepest basin in the ocean, we prove the following relation between total numbers of critical points of function $d$ on planet $\mathcal{P}$:

$$m - p + M = 2. \tag{4}$$

A slightly different exposition of formulae (3) and (4) can be found in the inspirational paper [3].

We are ready to prove the famous Euler formula for polyhedra (proved in more geometrical manner in [5]). Let $\Pi$ be a convex polyhedron with $T$ vertices, $I$ edges and $S$ faces. Suppose that there is an interior point $O$ of polyhedron $\Pi$ such that all perpendiculars from $O$ to any of faces end in interior points of faces. We may treat polyhedron $\Pi$ as some planet $\mathcal{P}$ and construct a distance function $d$ from the point $O$. The distance function $d$ has minima at ends of perpendiculars to faces, saddles at ends of perpendiculars to edges and maxima at vertices. According to the formula (4) we get

$$T - I + S = 2. \tag{5}$$

The other interesting way to prove Euler formula for polyhedra are based on construction of discrete gradient field. Let $\Pi$ be a convex polyhedron and $\{F_1, \dots, F_S\}$ be the set of barycenters of its faces, $\{E_1, \dots, E_I\}$ be the set of centers of its edges and $\{V_1, \dots, V_T\}$ be the set of its vertices. To each vertex put vectors towards centers of edges and barycenters of faces that are incidence to the given vertex. Similarly, to centers of edges put vectors towards barycenters of incidence faces. We obtain a directed graph on the surface of the polyhedron $\Pi$ which can be simply extended to a gradient field. That gradient field has maxima at vertices, saddles at centers of edges and minima at barycenters of faces. By equation (4) we obtain again relation (5).

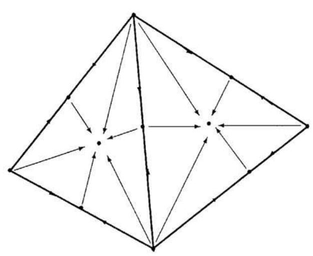


Fig. 4. A discrete gradient field

## 5. Conclusion

Morse theory is a powerful tool for the study of surfaces and their higher dimensional analogues called manifolds. Manifolds are spaces that locally look like Euclidean spaces. Morse function on a manifold gives a way to describe how this manifold is built from simple parts called handles. Morse theory and its analysis of the evolution of level sets of Morse functions provides a rich source of ideas to apply topology in many mathematical problems.

I am indebted to Professor Milosav Marjanović who introduced me to this beautiful and powerful theory.

**REFERENCES**

[1] J.Milnor, *Morse Theory*, Princeton University Press, 1963.

[2] J.C. Maxwell, *On Hills and Dales*, The Philosophical Magazine **40** (269) (1870), 421–427, http://en.m.wikipedia.org/wiki/Morse_theory.

[3] М. Шубин, *Топология и рельеф местности*, Квант **8** (1982), 10--15.

[4] V. Janković, *Quadratic functions in several variables*, The Teaching of Mathematics **VIII** (2) (2005), 53–60.

[5] S. Vrećica, *On polygons and polyhedra*, The Teaching of Mathematics **VIII** (1) (2005), 1–14.

University of Belgrade, Faculty of Mathematics, Studentski trg 16, 11000 Beograd, Serbia

*E-mail*: vgrujic@matf.bg.ac.rs