

# *Morse Theory, Discrete Morse Theory and Applications*

Zexuan Dong<sup>1,a,\*</sup>

<sup>1</sup>Shanghai Pinghe School, Shanghai, China

a. Kzm9939@outlook.com

\*corresponding author

**Abstract:** By employing a specific class of smooth functions to study a space, Morse theory establishes deep connections between analysis and topology. It is a classical subject of pure mathematics, originally pioneered by Marston Morse in the 1920s. In this article, we use Morse theory to present a proof of an interesting result on the knots, known as the Fáry-Milnor theorem. We also discuss discrete Morse theory, a subject of applied mathematics developed by Robin Forman in the 1990s, and its application. We focus on elucidating especially the inherent similarity between classical Morse theory and discrete Morse theory.

**Keywords:** Topology, Morse functions, knots, simplicial homology.

## 1. Introduction

Consider a knot, which is a one-dimensional object in the space, twisted in an arbitrary way to form a closed loop, like a shoelace, but with both ends connected. Mathematically, a **knot** can be understood as a continuous and injective map  $\gamma: S^1 \rightarrow \mathbf{R}^3$ . Since  $S^1$  is compact,  $\gamma(S^1)$  has the homeomorphism type of  $S^1$  itself. So the intrinsic topological structure of a knot  $K$  is always merely a circle, but the complexity of the way  $K$  wraps around is encoded in the map  $\gamma$ . The simplest example of  $\gamma$  is just a plain embedding, e.g. it (denoted as  $\gamma_0$ ) embeds  $S^1$  as

$$\{(x, y, z) \in \mathbf{R}^3 | x^2 + y^2 = 1, z = 0\} \quad (1)$$

Such a knot is completely without entanglement. When  $\gamma$  as a map becomes complicated, the knot  $K$  might be entangled "in an essential way". The following formal definition of the **isotopy** relation between knots captures the intuitive process of deforming a knot.

**Definition 1.1.** An isotopy from a knot  $(K_1; \gamma_1)$  to another knot  $(K_2; \gamma_2)$  is a homotopy  $f: S^1 \times [0, 1] \rightarrow \mathbf{R}^3$  from  $\gamma_1$  to  $\gamma_2$  (meaning a continuous map with  $f(-, 0) = \gamma_1$ ,  $f(-, 1) = \gamma_2$ ), such that  $\forall t \in [0, 1]$ ,  $f(-, t)$  is always a knot, in the sense mentioned above. If such an  $f$  exists, we say that  $K_1$  and  $K_2$  are isotopic, or equivalent.

**Definition 1.2.** Any knot equivalent to  $\gamma_0$  defined by 1.1 is said to be an **unknot**, or can be unknotted.

Sometimes in practice, additional structure needs to be put on a knot to study it, namely the smooth structure. So we view  $\gamma$  as a smooth map between the smooth manifolds  $S^1$  and  $\mathbf{R}^3$ , hence we can speak of the tangent map  $\gamma'$ . Since the shape of the knot is determined only by the image of  $\gamma$ , we can choose any reparameterization. We can always choose one that satisfies the uniformization

convention: the tangent vectors are unit vectors everywhere (called the arc length reparameterization). To summarize, we now revise our definition of knots as follows.

**Definition 1.3.** A knot  $K$  is a smooth embedding  $\gamma$  from the quotient manifold  $\mathbf{R}/L \cong S^1$  to  $\mathbf{R}^3$  (i.e. a smooth map  $\gamma = (\gamma_1, \gamma_2, \gamma_3): \mathbf{R} \rightarrow \mathbf{R}^3$  with  $\gamma(t) = \gamma(t + L)$ ), such that  $|\gamma'(t)| = 1, \forall t \in \mathbf{R}$ . The number  $L$  is the length of  $K$ .

**Definition 1.4.** The local curvature of  $K$  is a function defined as

$$\kappa(t) = |\gamma''(t)|, \forall t \in \mathbf{R}. \quad (2)$$

The **total curvature** of  $K$  is a number that is the integration of local curvatures:

$$T_K := \int_0^L \kappa(t) dt. \quad (3)$$

The local curvature reflects how curved the knot is locally at a point  $\gamma(t)$ :  $\kappa(t)$  is large when bending very sharply, small when barely bending (and 0 when it is locally just a flat line). Imagine the knot as the trajectory of a uniform motion, then the curvature represents the magnitude of acceleration. The larger it is, the faster the point deviates. If we integrate all the local curvatures together, we obtain a quantity that measures the overall twistedness of the knot. The following result (to be proved) is a quantitative way of saying that for a knot to be linked head-to-tail, it has to be "sufficiently curved".

**Theorem 1.5.**  $T_K \geq 2\pi$ , for any knot  $K$ .

The next result, originally proved by Fáry [1] and Milnor [2] and reviewed in [3], is a central problem that we are going to address in this article. Intuitively, it says that if a knot is not "curved enough", then it can be unknotted.

**Theorem 1.6.** If  $T_K < 4\pi$ , then  $K$  is an unknot.

**Example.** Take the trefoil knot for example. It can be wrapped around a torus as below:

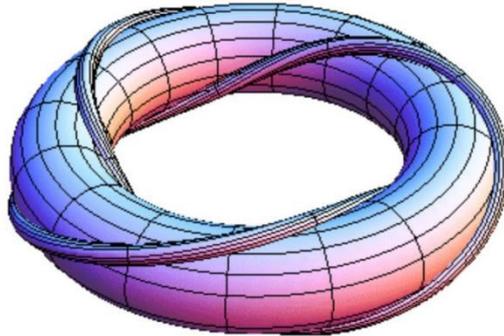


Figure 1: Picture taken from <https://mathoverflow.net/questions/91444>.

It is essentially entangled (can not be unknotted, see [4]). The torus surface can be defined by the equation in terms of  $(r \cos \theta, r \sin \theta, z)$ :

$$(r - 2)^2 + z^2 = 1. \quad (4)$$

Observing how the trefoil knot is wrapped on that surface, we can give it a parameterization:

$$\gamma(t) = \left( (2 + \cos 3t) \cos 2t, (2 + \cos 3t) \sin 2t, \sin 3t \right), t \in [0, 2\pi]. \quad (5)$$

It is not an arc length parameterization. Nevertheless there is still a formula computing the curvature ([5] theorem 7.15):

$$\kappa = \frac{|\gamma' \times \gamma''|}{|\gamma'|^3} \quad (6)$$

where  $\times$  is the cross product. Plugging in the formula of  $\gamma$ , a software computation reveals that

$$T = \int_0^{2\pi} \kappa(t) dt > 4\pi. \quad (7)$$

This does not violate the Fáry-Milnor theorem. Note that the theorem states that  $T_K < 4\pi$  implies unknot, but does not say anything when  $T_K \geq 4\pi$ . In fact, no matter how we deform the trefoil knot, the total curvature can not be less than  $4\pi$ .

**Remark 1.7.** We emphasize that whether a knot is an unknot or not is a topological statement. It is a well-defined property that does not depend on which specific smooth structure / parameterization is given to the knot. So in the Fáry-Milnor theorem, we are studying this topological property of a knot through the lens of smooth data on it, i.e. the curvatures.

The method that we are going to employ to prove the theorem, besides basic vector calculus, is the **Morse theory**, which does not show up explicitly in the original papers of Fáry-Milnor. It is a rich and beautiful theory in its own right. We will give a quick review of its basics in section 2. In section 3 we illustrate this theory with a simple and non trivial example  $\mathbf{RP}^n$ . We find this example very much exhibits the essence of Morse theory. After these preparatory works, we present in detail a proof of the Fáry-Milnor theorem in section 4. In the last section we study the discrete Morse theory. The key is to elucidate how the concepts in classical Morse theory are parallelly borrowed in the setting of discrete world. Then as an application, we discuss how the discrete Morse theory is related to the computation of simplicial homology.

## 2. Review of Morse theory

We state the basic results in classical Morse theory. The standard reference is [6] or [7]. First we need some definitions. Throughout,  $M$  is a compact smooth manifold of dimension  $n$ , and  $f \in C^\infty(M)$  is a smooth function on it.

**Definition 2.1** For  $x \in M$ , let  $\nabla_x f$  denote the induced linear map on the tangent spaces

$$\nabla_x f: T_x M \rightarrow T_{f(x)} \mathbf{R} = \mathbf{R}. \quad (8)$$

In a local coordinate, we can represent the map as a  $1 \times n$  matrix:

$$\nabla_x f = \left( \frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right). \quad (9)$$

This is called the **gradient** of  $f$  at  $x$ .

Fix an embedding into Euclidean space of  $M$  (which is always possible by Whitney's theorem, see [8] theorem 6.15). Often by abus of notation, we understand  $\nabla_x f$  as an element in  $T_x M$  (i.e. a tangent vector):

$$\nabla_x f = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x) \cdot \frac{\partial}{\partial x_i} \quad (10)$$

where  $\left\{ \frac{\partial}{\partial x_i} \right\}$  is chosen to be an orthonormal basis of  $T_x M$ . This is well-defined (does not depend on the choice of the basis  $\left\{ \frac{\partial}{\partial x_i} \right\}$ ), and represents the vector of steepest ascent of  $f$  at  $x$ .

**Definition 2.2** A **critical point** of  $f$  is a point  $p \in M$  such that  $\nabla_p f = 0$ .

Concretely, this means that  $f$  has 0 rate of change along any direction at  $p$ .

**Morse lemma.** At a critical point  $p$ , there is a local coordinate chart  $(x_i)$  (called a standard coordinate chart) centered at  $p$  (meaning  $p = (0, \dots, 0)$ ) in which  $f$  can be locally expressed as

$$f(x) = -x_1^2 - \dots - x_\lambda^2 + x_{\lambda+1}^2 + \dots + x_\mu^2. \quad (11)$$

Moreover, the numbers  $0 \leq \lambda \leq \mu \leq n$  are uniquely determined by  $f$  and the point  $p$ . We say that  $\lambda$  is the **index** of  $p$ .

As a consequence, critical points are isolated hence finitely many (since  $M$  is compact).

**Definition 2.3.** A critical point is said to be **non degenerate** if  $\mu = n$ , and  $f$  is said to be a **Morse function** if all critical points of  $f$  are non degenerate.

As a smooth vector field,  $\nabla f$  determines a flow called the gradient flow  $X(t)$  (a smooth curve on  $M$ ) in the sense that

$$X'(t) = \nabla_{X(t)} f. \quad (12)$$

Such a curve / flow always exists locally due to the theorem of existence and uniqueness of solutions of ordinary differential equations (known as the Picard-Lindelöf theorem, [9] theorem 2.2). If we consider the gradient flow determined by  $-\frac{\nabla f}{|\nabla f|^2}$  (i.e. the same flow curve but with a reparameterization), then ([6] theorem 3.1) we have a uniform descent and it turns out that:

**The 1st fundamental theorem of Morse theory.** Denote  $M_c := f^{-1}(-\infty, c]$ ,  $c \in \mathbf{R}$ . If  $a < b$  and  $f^{-1}[a, b]$  contains no critical point, then we have a diffeomorphism  $M_a \cong M_b$ .

In the case of the presence of critical points, we have another result (theorem 3.2):

**The 2nd fundamental theorem of Morse theory.** If  $f^{-1}[a, b]$  contains one critical point  $p$  and it is non degenerate of index  $\lambda$ , then  $M(b)$  deformation retracts onto  $M(a)$  with a  $\lambda$ -dimensional cell  $e^\lambda$  attached (at the boundary), so that we have a homotopy equivalence  $M(b) \simeq M(a) \cup e^\lambda$ .

Put together, we conclude that if  $f$  is a Morse function with  $k$  critical points of indices  $\lambda_1 \leq \dots \leq \lambda_k$  respectively, then  $M$  has the homotopy type  $e^{\lambda_1} \cup \dots \cup e^{\lambda_k}$  of a cell complex (see the appendix of [10]) of exactly  $k$  cells, one for each critical point of  $f$ .

**Remark 2.4.** A subtle issue arises when two critical points happen to have the same value. But it turns out that one can always perturb  $f$  slightly so that the values are distinct (while preserving the indices), then the conclusion follows from the two fundamental theorems of Morse theory.

### 3. Morse theory on the real projective space $\mathbf{RP}^n$

In this section we work out explicitly the Morse theory when applied to  $\mathbf{RP}^n$ , the real projective space. Though there is no logical dependence for the subsequent sections, we found this explicit example very helpful during the author's studying process of Morse theory.

#### 3.1. $\mathbf{RP}^n$ as a manifold

The space  $\mathbf{RP}^n$  is defined as consisting of all lines in  $\mathbf{R}^{n+1}$  passing through the origin. As a quotient space,

$$\mathbf{RP}^n = \{(x_0, x_1, \dots, x_n) \in \mathbf{R}^{n+1} - \{0\} \mid (x_0, \dots, x_n) \sim (\lambda x_0, \dots, \lambda x_n), \lambda \in \mathbf{R} - \{0\}\} \quad (13)$$

or equivalently

$$\mathbf{RP}^n = S^n / \{\pm I\} = \{(x_0, \dots, x_n) \in \mathbf{R}^{n+1} \mid \sum_{i=0}^n x_i^2 = 1\} / (x_0, \dots, x_n) \sim (-x_0, \dots, -x_n). \quad (14)$$

**Proposition 3.1.**  $\mathbf{RP}^n$  is a compact smooth manifold of dimension  $n$ .

Proof: Since  $S^n$  is compact,  $\mathbf{RP}^n$  is compact as its quotient space. To show the smoothness, we construct a coordinate atlas as follows. In 3.1 we denote the equivalence class of  $(x_0, \dots, x_n)$  as  $[x_0, \dots, x_n]$  (the homogeneous coordinate). Notice that  $\mathbf{RP}^n$  is covered by  $n + 1$  open subsets

$$U_i := \{[x_0, x_1, \dots, x_n] | x_i \neq 0\}, \quad i = 0, 1, \dots, n. \quad (15)$$

On each  $U_i$ , we have a well-defined continuous map  $\varphi_i: U_i \rightarrow \mathbf{R}^n$  as

$$\varphi_i([x_0, x_1, \dots, x_n]) = \left(\frac{x_0}{x_i}, \dots, \frac{x_{i-1}}{x_i}, \frac{x_{i+1}}{x_i}, \dots, \frac{x_n}{x_i}\right) \quad (16)$$

with a continuous inverse

$$\varphi_i^{-1}(a_1, \dots, a_n) = [a_1, \dots, a_i, 1, a_{i+1}, \dots, a_n]. \quad (17)$$

It follows that  $\varphi_i$  is a homeomorphism, hence defines a coordinate chart of  $U_i$ . For  $i < j$ , on  $U_i \cap U_j = \{[x_0, x_1, \dots, x_n] | x_i, x_j \neq 0\}$ , the coordinate transformation

$$\varphi_j \circ \varphi_i^{-1}(a_1, \dots, a_n) = \left(\frac{a_1}{a_j}, \dots, \frac{a_i}{a_j}, \frac{1}{a_j}, \dots, \frac{a_{j-1}}{a_j}, \frac{a_{j+1}}{a_j}, \dots, \frac{a_n}{a_j}\right) \quad (18)$$

is a smooth map. We conclude that  $\{U_i; \varphi_i\}_{i=0,1,\dots,n}$  is a smooth atlas, and  $\mathbf{RP}^n$  is a smooth manifold of dimension  $n$ .  $\square$

### 3.2. A Morse function on $\mathbf{RP}^n$

In this subsection we construct a function on  $\mathbf{RP}^n$  and show that it is a Morse function.

**The construction of the function.** Take an arbitrary sequence of numbers  $c_0 < c_1 < \dots < c_n$ . Consider the smooth function

$$\tilde{f}(x_0, x_1, \dots, x_n) = \sum_{i=0}^n c_i x_i^2 \quad (19)$$

on  $\mathbf{R}^{n+1}$ . It restricts to a smooth function  $\tilde{f}|_{S^n}$  on the unit sphere  $S^n$ , which then descends to a function  $f$  on  $\mathbf{RP}^n$  via the identification [sn]. That is,

$$\tilde{f}|_{S^n} = f \circ q \quad (20)$$

where  $q$  denotes the quotient map  $S^n \rightarrow S^n / \{\pm 1\} = \mathbf{RP}^n$ .

**Proposition 3.2.**  $f$  is a well-defined smooth function.

Proof: It is well-defined since

$$\tilde{f}(x_0, x_1, \dots, x_n) = \tilde{f}(-x_0, -x_1, \dots, -x_n). \quad (21)$$

As a polynomial, it is clearly smooth.  $\square$

We can easily compute the gradient of  $\tilde{f}$ :

$$\nabla_x \tilde{f} = \left(\frac{\partial \tilde{f}}{\partial x_0}(x), \dots, \frac{\partial \tilde{f}}{\partial x_n}(x)\right) = 2(c_0 x_0, \dots, c_n x_n), \quad x = (x_0, \dots, x_n) \in \mathbf{R}^{n+1}. \quad (22)$$

Since the tangent space of  $S^n$  is included into that of  $\mathbf{R}^{n+1}$  as

$$T_x S^n = \{(y_0, \dots, y_n) | (y_0, \dots, y_n) \perp (x_0, \dots, x_n)\} \hookrightarrow T_x \mathbf{R}^{n+1} = \mathbf{R}^{n+1} \quad (23)$$

the tangent map induced by  $\tilde{f}|_{S^n}$  is

$$\nabla_x \tilde{f}|_{S^n} \cdot T_x S^n \rightarrow \mathbf{R}, (y_0, \dots, y_n) \mapsto \sum_{i=0}^n 2c_i x_i y_i = 2(c_0 x_0 \dots, c_n x_n) \cdot (y_0 \dots, y_n). \quad (24)$$

By the chain rule and 3.8 we have

$$\nabla_x \tilde{f}|_{S^n} = \nabla_x f \circ q_{*,x} \quad (25)$$

where the tangent map  $q_{*,x}: T_x S^n \rightarrow T_x \mathbf{RP}^n$  is an isomorphism. We deduce that

$$\nabla_x f = 0 \Leftrightarrow \nabla_x \tilde{f}|_{S^n} = 0 \Leftrightarrow (c_0 x_0, \dots, c_n x_n) \in (T_x S^n)^\perp = \mathbf{R} \cdot (x_0, \dots, x_n). \quad (26)$$

But the  $c_i$ 's are distinct to each other, we have  $(c_0 x_0, \dots, c_n x_n)$  is colinear with  $(x_0, \dots, x_n) \Leftrightarrow$  all but one  $x_i$  are zero. In conclusion, we have shown that

**Proposition 3.3.** The critical points of  $f$  are

$$p_0 := [1, 0, \dots, 0], p_1 := [0, 1, 0, \dots, 0], \dots, p_n := [0, \dots, 0, 1] \quad (27)$$

To show that  $f$  is a Morse function, it remains to show that each  $p_i$  is non degenerate. In fact

**Proposition 3.4.**  $p_i$  is non degenerate with index  $i$ .

Proof: Consider the open subset  $V_i \subset S^n$  defined by

$$V_i = \{(x_0, \dots, x_n) \in S^n | x_i > 0\}. \quad (28)$$

Then the projection map

$$\psi_i: V_i \rightarrow \mathbf{R}^n, (x_0, \dots, x_n) \mapsto (x_0, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \quad (29)$$

is a homeomorphism onto the image of  $\psi_i$ , thus makes  $V_i$  into a local coordinate chart centered at  $p_i$ . On this chart we can write  $f$  as

$$f|_{V_i} \circ \psi_i^{-1}: (x_0, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \mapsto \sum_{i=0}^n c_i x_i^2 = c_i + \sum_{j \neq i} (c_j - c_i) x_j^2 \quad (30)$$

where we have used the relation

$$x_i = \sqrt{1 - \sum_{j \neq i} x_j^2}. \quad (31)$$

Since  $c_j - c_i < 0$  for  $j < i$ , and  $c_j - c_i > 0$  for  $j > i$ , locally at  $p_i$ ,  $f$  has exactly  $i$  negative directions and  $n - i$  positive directions. A dilation transformation

$$(x_0, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \mapsto \left( \frac{x_0}{\sqrt{c_i - c_0}}, \dots, \frac{x_{i-1}}{\sqrt{c_i - c_{i-1}}}, \frac{x_{i+1}}{\sqrt{c_{i+1} - c_i}}, \dots, \frac{x_n}{\sqrt{c_n - c_i}} \right) \quad (32)$$

takes  $f$  into a standard form as in the Morse lemma. We thus conclude that  $p_i$  is non degenerate with index  $i$ .  $\square$

### 3.3. Applying the Morse theory

Now we can inspect what the Morse theory states for this particular Morse function  $f$  on  $\mathbf{RP}^n$ . By applying the two fundamental theorems of Morse theory we arrive at:

**Proposition 3.5.**  $\mathbf{RP}^n$  has the homotopy type  $e^0 \cup e^1 \cup \dots \cup e^n$ .

In fact, we know that this is exactly the cell decomposition of  $\mathbf{RP}^n$ . Consider the equator  $S^{n-1}$  in  $S^n$ , then

$$\mathbf{RP}^{n-1} = S^{n-1}/\{\pm 1\} \subset \mathbf{RP}^n = S^n/\{\pm 1\} \quad (33)$$

and  $\mathbf{RP}^n - \mathbf{RP}^{n-1}$  can be identified with a hemisphere that is, an  $n$ -dimensional cell. The upshot is that the Morse function  $f$ , which is a smooth datum, indeed provides us with some useful topological information about  $\mathbf{RP}^n$ .

#### 4. Proving the Fáry-Milnor theorem

We are now ready to prove theorem 1.5 and 1.6. To use Morse theory, we will employ height functions to supply Morse functions on  $K$ . Precisely, for any  $v \in S^2$  the unit sphere of  $\mathbf{R}^3$ , consider the height function  $h_v$  on  $K$  along  $v$

$$h_v(x) := \langle x, v \rangle \quad (34)$$

where  $\langle -, - \rangle$  is the inner product on  $\mathbf{R}^3$ . Denote by  $\mu(v)$  the number of critical points of  $h_v$ . The key in our proof is to show that the total curvature naturally emerges in the averaging process of the numbers  $\mu(v)$ :

$$\frac{T_K}{\pi} = \frac{1}{4\pi} \int_{S^2} \mu(v) dv. \quad (35)$$

Then we will subsequently show that

**Proposition 4.1.** If  $T_K < 4\pi$ , there exists an element  $v \in S^2$  such that  $\mu(v) = 2$ .

**Proposition 4.2.** If  $h_v$  has exactly two critical points, then  $K$  can be unknotted.

We start with deriving the equation 35.

##### 4.1. The Morse functions $h_v$

First of all, not all height functions  $h_v$  are Morse functions. We will characterize the condition on  $v$  such that  $h_v$  is not a Morse function. On such points  $v \in S^2$ ,  $\mu(v)$  is ill-defined. However, we will see that such points form a measure zero subset of  $S^2$ , hence do not affect the integral in 35.

We first characterize the condition of  $\gamma(s) \in K$  being a critical point of  $h_v$ . In the local coordinate  $(s)$  of  $K$ , since  $h_v(s) = \langle \gamma(s), v \rangle$ , we have

$$\nabla h_v = h'_v(s) = \langle \gamma'(s), v \rangle. \quad (36)$$

The gradient at  $\gamma(s)$  vanishes if and only if  $v \perp \gamma'(s)$ . This motivates the following definition.

**Definition 4.3.** Let  $S(K)$  be a surface defined as

$$S(K) := \{(\gamma(s), v) \in K \times S^2 \mid v \perp \gamma'(s), s \in \mathbf{R}\}. \quad (37)$$

**Proposition 4.4.**  $S(K)$  is a smooth manifold of dimension 2.

Proof: Recall that  $\gamma'(s)$  is a unit vector. Consider the open submanifold of  $K \times \mathbf{R}^3$  defined as

$$\tilde{S}(K) := \{(\gamma(s), v) \in K \times \mathbf{R}^3 \mid v \text{ is not colinear with } \gamma'(s)\}. \quad (38)$$

Let  $f: \tilde{S}(K) \rightarrow \mathbf{R}^2$  be a smooth map defined as

$$f: (\gamma(s), v) \mapsto (v \cdot \gamma'(s), v \cdot v) \quad (39)$$

where denotes the inner product. Since its Jacobian matrix

$$J(f) = \begin{pmatrix} v \cdot \gamma''(s) & \gamma'(s) \\ 0 & 2v \end{pmatrix}$$

is of full rank everywhere ( $2v$  is not colinear with  $\gamma'(s)$ ), by the constant rank theorem ( theorem 5.12),  $S(K) = f^{-1}(0,1)$  is a regular submanifold of  $\tilde{S}(K)$  of codimension 2. Thus  $S(K)$  is a smooth manifold of dimension 2. ◻

**Remark 4.5.** For a given  $\gamma(s)$  on the knot, the points  $v$  such that  $v \perp \gamma'(s)$  form a great circle of  $S^2$ . In formal language, we say that  $S(K)$  is the normal circle bundle on  $K$ . Intuitively, we can think of  $S(K)$  as the surface of a rope representing  $K$  with thickness. This is justified by the following result.

**Proposition 4.6.**  $S(K)$  can be embedded into  $\mathbf{R}^3$  via

$$S(K) \hookrightarrow \mathbf{R}^3, (\gamma(s), v) \mapsto \gamma(s) + \epsilon v \quad (40)$$

for a sufficient small constant  $\epsilon > 0$ .

Proof: This directly follows from the tubular neighborhood theorem ( theorem 6.24). Alternatively, we can sketch a proof without invoking the theorem. By the compactness of  $K$ , let  $M$  be the maximum of the local curvatures  $\kappa(s) = |\gamma''(s)|$ . We can choose  $\epsilon \in (0, \frac{1}{2M})$  such that every ball of radius  $\epsilon$  intersects  $K$  at a connected interval of length  $< \frac{1}{M}$  (again by the compactness). We show that  $\exists \gamma(s_1) + \epsilon v_1 = \gamma(s_2) + \epsilon v_2$  for  $s_1 < s_2$ , so that  $\epsilon$  satisfies the requirement. If such  $s_1$  and  $s_2$  exist, then  $s_2 - s_1 < \frac{1}{M}$ . Thus

$$|\gamma'(s) - \gamma'(s_1)| \leq (s - s_1)M < 1, \forall s \in [s_1, s_2]. \quad (41)$$

That is,  $\gamma'(s), \gamma'(s_1) \in S^2$  have central angle  $< \frac{\pi}{3}$ . This implies that the distance from  $\gamma(s_2)$  to the normal plane  $\mathcal{P}_1$  at  $\gamma(s_1)$  is at least  $(s_2 - s_1)\cos\frac{\pi}{3} = \frac{s_2 - s_1}{2}$ . Moreover, if  $\alpha$  is the dihedral angle between the normal plane  $\mathcal{P}_2$  at  $\gamma(s_2)$  and  $\mathcal{P}_1$ , then  $\sin\alpha < (s_2 - s_1)M$ . We conclude that the distance from  $\gamma(s_2)$  to  $\mathcal{P}_1 \cap \mathcal{P}_2$  is at least  $\frac{s_2 - s_1}{2\sin\alpha} > \frac{1}{2M} > \epsilon$ , a contradiction. ◻

From now on, we will make no distinction between  $S(K)$  and the embedded manifold, and  $\epsilon$  will be suppressed. The embedding enables us to speak of  $T_x S(K)$  as an inner product space. Along the great circle, we have a tangent vector  $\frac{\partial}{\partial \theta} \in T_x S(K)$ . Perpendicular to it, we have  $\frac{\partial}{\partial s}$  ("along the knot") so that  $\{\frac{\partial}{\partial s}, \frac{\partial}{\partial \theta}\}$  forms an orthonormal basis of  $T_x S(K)$ .

**Remark 4.7.** To motivate the notation  $\frac{\partial}{\partial s}$ , notice that the tangent plane  $T_x S(K)$  is parallel to  $\gamma'(s)$  (where  $x = (\gamma(s), v)$ ), hence  $\frac{\partial}{\partial s}$  can be viewed as the same vector as  $\gamma'(s)$ . To see this, consider an arbitrary curve  $\gamma(s) + v(s)$  on  $S(K)$ . We have

$$\langle \gamma'(s) + v'(s), v(s) \rangle = \langle \gamma'(s), v(s) \rangle + \langle v'(s), v(s) \rangle = 0 \quad (42)$$

where  $\langle \gamma'(s), v(s) \rangle = 0$  follows from the definition of  $S(K)$ ,  $\langle v'(s), v(s) \rangle = 0$  follows from  $|v(s)| \equiv 1$ . So all tangent vectors are perpendicular to  $v(s)$ , hence the tangent plane is parallel to  $\gamma'(s)$ .

Consider the natural projection

$$p: S(K) \rightarrow S^2, (\gamma(s), v) \mapsto v. \quad (42)$$

**Proposition 4.8.** If  $h_v$  is not a Morse function, then  $v$  is a critical value of  $p$ .

Proof: If so, there is a degenerate critical point  $\gamma(s_0)$ . Recall that  $\gamma(s_0)$  being a critical point means  $x := (\gamma(s_0), v) \in S(K)$ . We show that  $p_*: T_x S(K) \rightarrow T_v S^2$  is not surjective, hence  $v$  is a critical value. The  $\gamma(s_0)$  being degenerate means the Hessian matrix of  $h_v$  at it is singular ([6], section 1.2).

In our case, the Hessian matrix is just the  $1 \times 1$  matrix  $h_v''(s_0) = \gamma''(s_0) \cdot v$  so it is zero. For any curve  $(\gamma(s), v(s))$  passing through  $x$  on  $S(K)$ , since  $\gamma'(s) \cdot v(s) \equiv 0$ , taking derivative at  $s_0$  yields

$$0 = \gamma''(s_0) \cdot v + \gamma'(s_0) \cdot v'(s_0) = \gamma'(s_0) \cdot v'(s_0). \quad (43)$$

That is,  $v'(s_0) \in \gamma'(s_0)^\perp$  which is a 1-dimensional subspace of  $T_v S^2$ . Hence  $p_*$  is not surjective (in fact  $p_*$  is of rank 1 with image generated by  $p_* \left( \frac{\partial}{\partial \theta} \right)$ ).  $\square$

**Corollary 4.9.** The points  $v \in S^2$  such that  $h_v$  is not Morse form a measure zero subset.

Proof: This follows from Sard's theorem ([8] theorem 6.10).  $\square$

## 4.2. Evaluating the integral

We just proved that the integral in 35 is well-defined, now we evaluate it. Recall that  $\mu(v)$  as the number of critical points of  $h_v$  can be expressed as

$$\mu(v) = |\{\gamma(s) \in K \mid (\gamma(s), v) \in S(K)\}|. \quad (44)$$

Hence  $\mu(v)$  is exactly the cardinality of the preimage  $p^{-1}(v)$ .

**Proposition 4.10.** At each point  $x \in S(K)$  we define  $x_s := p_* \left( \frac{\partial}{\partial s} \right)$ ,  $x_\theta := p_* \left( \frac{\partial}{\partial \theta} \right)$ , and a matrix

$$G(x) := \begin{pmatrix} \langle x_s, x_s \rangle & \langle x_s, x_\theta \rangle \\ \langle x_\theta, x_s \rangle & \langle x_\theta, x_\theta \rangle \end{pmatrix}. \quad (45)$$

Then we have

$$\int_{S(K)} \sqrt{\det(G)} dsd\theta = \int_{S^2} \mu(v) dv. \quad (46)$$

Proof: This is the coarea formula ([11] theorem 3.11). However, we can sketch an informal heuristic argument. The matrix  $G(x)$  is so defined that  $\sqrt{\det(G(x))}$  is the local ratio induced by  $p$  at  $x$  of infinitesimal areas. Explicitly, we can compute the determinant as

$$\det(G(x)) = |x_s|^2 |x_\theta|^2 - (|x_s| |x_\theta| \cos \alpha)^2 = |x_s|^2 |x_\theta|^2 \sin^2 \alpha \quad (47)$$

where  $\alpha$  is the angle between  $x_s$  and  $x_\theta$  in  $T_{p(x)} S^2$ . This is exactly the squared area of the infinitesimal parallelogram spanned by  $x_s, x_\theta$ . Recall that  $\left\{ \frac{\partial}{\partial s}, \frac{\partial}{\partial \theta} \right\}$  is an orthonormal basis, we conclude that infinitesimally  $p$  scales up the area element  $dsd\theta$  of  $S(K)$  into  $S^2$  by the scalar  $\sqrt{\det(G)}$ . Integrating  $\sqrt{\det(G)} dsd\theta$  on  $S(K)$  is to integrate a weighted area element on  $S^2$ , where  $dv$  is weighted  $\mu(v) = |p^{-1}(v)|$  times since  $v$  is covered such many times under  $p$ .  $\square$

To evaluate  $\int_{S(K)} \sqrt{\det(G)} dsd\theta$ , we first integrate along  $\theta$  (i.e. along the great circles):

**Proposition 4.11.** For each  $s$ ,

$$\int_0^{2\pi} \sqrt{\det(G)} d\theta = 4\kappa(s). \quad (48)$$

This section will culminate at this central result. We will give two proofs, one is computational, and one is conceptual and informal. Note that equation 35 will be derived from this:

$$\frac{T_K}{\pi} = \frac{l}{\pi} \int_0^L \kappa(s) ds = \frac{l}{4\pi} \int_{S(K)} \sqrt{\det(G)} dsd\theta = \frac{l}{4\pi} \int_{S^2} \mu(v) dv. \quad (49)$$

Proof: (*Computational.*) To compute, we choose a local orthonormal moving frame  $\{\gamma'(s), e_1(s), e_2(s)\}$  along  $K$ , where  $e_1(s)$  is a flow curve on  $S(K)$  generated by  $\frac{\partial}{\partial s}$ , and  $e_2(s) = \gamma'(s) \times e_1(s)$ . We can think of  $e_1(s) / e_2(s)$  as marking the curve  $\theta = 0 / \theta = \frac{\pi}{2}$ . We can thus express the map  $p$  in terms of local coordinate

$$p(s, \theta) = \cos\theta e_1(s) + \sin\theta e_2(s). \quad (50)$$

It then follows that

$$x_s = \cos\theta e_1'(s) + \sin\theta e_2'(s), \quad x_\theta = -\sin\theta e_1(s) + \cos\theta e_2(s). \quad (51)$$

Recall that (remark4.7)  $\frac{\partial}{\partial s} = \gamma'(s)$  (as vectors in  $\mathbf{R}^3$ ), there is a function  $a$  in  $s$  such that

$$e_1' = a\gamma'. \quad (52)$$

Since  $e_2' \perp e_2$ ,  $e_1' \perp e_2$  and

$$0 = \dot{\theta} = \langle e_1, e_2 \rangle' = \langle e_1', e_2 \rangle + \langle e_2', e_1 \rangle = \langle e_2', e_1 \rangle \quad (53)$$

we similarly have a function  $b$  in  $s$  such that

$$e_2' = b\gamma'. \quad (54)$$

Then we can compute the entries of  $G$  as

$$\langle x_\theta, x_\theta \rangle = \sin^2\theta + \cos^2\theta = 1. \quad (55)$$

$$\langle x_s, x_\theta \rangle = \langle (a\cos\theta + b\sin\theta)\gamma', -\sin\theta e_1 + \cos\theta e_2 \rangle = 0. \quad (56)$$

$$\langle x_s, x_s \rangle = \left| (a\cos\theta + b\sin\theta)\gamma' \right|^2 = (a\cos\theta + b\sin\theta)^2. \quad (57)$$

Thus

$$\sqrt{\det(G)} = |a\cos\theta + b\sin\theta| = \sqrt{a^2 + b^2} |\sin(\theta + \varphi)| \quad (58)$$

where  $\varphi = \arctan \frac{a}{b}$ . Then

$$\int_0^{2\pi} \sqrt{\det(G)} d\theta = \sqrt{a^2 + b^2} \int_0^{2\pi} |\sin(\theta + \varphi)| d\theta = 4\sqrt{a^2 + b^2}. \quad (59)$$

It remains to prove that  $\sqrt{a^2 + b^2} = |\gamma''(s)|$ . Since

$$0 = \dot{\theta} = \langle \gamma', e_1 \rangle' = \langle \gamma'', e_1 \rangle + \langle \gamma', e_1' \rangle = \langle \gamma'', e_1 \rangle + a \quad (60)$$

$$0 = \dot{\theta} = \langle \gamma', e_2 \rangle' = \langle \gamma'', e_2 \rangle + \langle \gamma', e_2' \rangle = \langle \gamma'', e_2 \rangle + b \quad (61)$$

we conclude that  $\gamma'' = -ae_1 - be_2$  hence  $|\gamma''| = \sqrt{a^2 + b^2}$ .  $\square$

**Remark 4.12.** Actually there is a subtle point in the proof: to justify the local coordinate system  $(s, \theta)$ , we need to prove that once  $e_1(s)$  is chosen / the curve  $\theta = 0$  is marked, then for any constant

$\varphi$ , the curve  $\theta = \varphi$  is also a flow curve generated by  $\frac{\partial}{\partial s}$ . That is, the tangent vector of the curve  $\cos\varphi e_1(s) + \sin\varphi e_2(s)$  is colinear with  $\frac{\partial}{\partial s}$  / perpendicular to  $\frac{\partial}{\partial \theta}$ . To see this, we show that

$$\langle \cos\varphi e_1'(s) + \sin\varphi e_2'(s), \frac{\partial}{\partial \theta} \rangle = 0. \quad (62)$$

Since  $e_1(s)$  is a flow curve of  $\frac{\partial}{\partial s}$ ,  $\langle e_1'(s), \frac{\partial}{\partial \theta} \rangle = 0$ . It remains to show that

$$\langle e_2'(s), \frac{\partial}{\partial \theta} \rangle = \langle \gamma'' \times e_1 + \gamma' \times e_1', \frac{\partial}{\partial \theta} \rangle = \det(\gamma'', e_1, \frac{\partial}{\partial \theta}) + \det(\gamma', e_1', \frac{\partial}{\partial \theta}) = 0. \quad (63)$$

This follows from the facts that  $\gamma'', e_1, \frac{\partial}{\partial \theta}$  are coplanar and  $\gamma', e_1'$  are colinear.

Proof: (*Conceptual.*) The formula essentially boils down to this elementary fact: on  $S^2$  the area of a pair of spherical lunes in a dihedral angle  $\alpha$  is  $4\pi \times \frac{\alpha}{\pi} = 4\alpha$ . Recall that  $\sqrt{\det(G)}$  is the local area ratio. Consider the great circles  $C_s$  and  $C_{s+ds}$  perpendicular to  $\gamma'(s)$  and  $\gamma'(s+ds)$  respectively, and let  $ds \rightarrow 0$  so that their dihedral angle is  $|\gamma''(s)|ds = \kappa(s)ds$ . Then  $\left(\int_0^{2\pi} \sqrt{\det(G)} d\theta\right) ds$  computes the sphere lunes area bounded by  $C_s$  and  $C_{s+ds}$ , which is  $4\kappa(s)ds$ . It follows that  $\int_0^{2\pi} \sqrt{\det(G)} d\theta = 4\kappa(s)$ .

### 4.3. The unknotting process

We will finally complete the proof. First notice a simple fact.

**Proposition 4.13.**  $\mu(v)$  is always an even number.

Proof: Suppose that  $h_v$  has  $n_0$  critical points of index 0,  $n_1$  critical points of index 1, then  $K \cong S^1$  has the homotopy type of a cell complex with  $n_0$  0-cells and  $n_1$  1-cells, thus its Euler number is  $n_0 - n_1$ . But  $\chi(S^1) = 0$ , so  $n_0 = n_1$ , hence  $\mu(v) = n_0 + n_1$  is even.

Alternatively, we have a proof without invoking the Euler number. By the Morse lemma, at the index-0 (resp. index-1) critical points  $h_v$  has local standard form  $x^2$  (resp.  $-x^2$ ), hence attains local minimum (resp. maximum). Conversely, a local extremum  $\gamma(s)$  of  $h_v$  must satisfy  $h_v'(s) = 0$ . Thus critical points are the same as local extrema. Since  $h_v$  is monotonic between two successive local extrema, the minima and the maxima must appear alternatively along the knot. We conclude that the number  $\mu(v)$  of local extrema is even.  $\square$

Moreover, as  $K$  is compact there is at least a global maximum and a global minimum, thus  $\mu(v) \geq 2$ . Together with equation 35 we have immediate corollaries:

Proof: (*of theorem 1.5.*) Since  $\frac{T_K}{\pi}$  is the average number of critical points, and the numbers are at least 2, we deduce that  $\frac{T_K}{\pi} \geq 2$  so  $T_K \geq 2\pi$ .  $\square$

Proof: (*of proposition 4.1.*) If  $T_K < 4\pi$ , then the average number  $\frac{T_K}{\pi} < 4$  hence  $\exists v \in S^2$  such that  $\mu(v) < 4$ . Since  $\mu(v)$  is a positive even number,  $\mu(v) = 2$ .  $\square$

To prove theorem 1.6, it remains to prove proposition 4.2. Now we have a given  $h_v$  with no local extrema except the global ones. Suppose  $h_v(K) = [a, b]$  and  $h_v(A) = a$ ,  $h_v(B) = b$  (i.e.  $A/B$  is the global minimum / maximum on  $K$ ). The complement of  $\{A, B\}$  in  $K$  has two components, which we denote as  $A \rightarrow B$  and  $B \rightarrow A$ . In the rest of the section, we describe an unknotting process of  $K$ .

Note now we are concerned with a purely topological statement as opposed to a smooth one. Recall that a topological knot is a continuous embedding  $\gamma: S^1 \hookrightarrow \mathbf{R}^3$ . To unknot  $K$ , our initial idea is as follows:  $\forall c \in (a, b)$ , consider the level plane  $h_v^{-1}(c)$  in  $\mathbf{R}^3$ . Since  $h_v$  on  $K$  between  $A$  and  $B$  is monotonic, the plane intersects  $A \rightarrow B$  and  $B \rightarrow A$  respectively at a unique point  $u_1$  and  $u_2$

(implicitly as functions in  $c$ ). Take any standard circle  $K_0$  with diameter  $AB$ , we want to show that  $K$  is isotopic to  $K_0$ , hence is an unknot. The tentative idea is to again take the intersection of the plane  $h_v^{-1}(c)$  with  $A \rightarrow B$  and  $B \rightarrow A$  in  $K_0$ , denoted as  $v_1$  and  $v_2$ , and consider a linear homotopy  $u_1 \rightarrow v_1$ ,  $u_2 \rightarrow v_2$  simultaneously in each level plane. However, the line segment  $u_1v_1$  and  $u_2v_2$  might intersect, which might then invalidate the isotopy. Explicitly, if  $\overrightarrow{u_1u_2} = \lambda \overrightarrow{v_2v_1}$  for some  $\lambda > 0$ , then at the moment  $t = \frac{\lambda}{\lambda+1}$  the two linear homotopies  $u_1 \rightarrow v_1$  and  $u_2 \rightarrow v_2$  crash each other. However, if instead  $\overrightarrow{u_1u_2} = \lambda \overrightarrow{v_1v_2}$  for some  $\lambda > 0$ , then  $u_1v_1$  will not intersect  $u_2v_2$ . Note that the direction of  $\overrightarrow{v_1v_2}$  is constant. This motivates the following final proof to remedy the issue just discussed.

Proof: (of theorem 1.6.) Consider the continuous map  $f$  from  $(a, b)$  to  $S^1$  defined as

$$f(c) = \frac{\overrightarrow{u_1u_2}}{|\overrightarrow{u_1u_2}|}. \quad (64)$$

Since  $(a, b)$  is contractible, any map from  $(a, b)$  is nullhomotopic. Take a homotopy  $F: (a, b) \times [0, 1] \rightarrow S^1$  from  $f$  to the constant map  $\frac{\overrightarrow{v_1v_2}}{|\overrightarrow{v_1v_2}|}$ . Then perform an isotopy of  $K$  as follows: in each level plane, rotate  $u_2$  around  $u_1$  along the path determined by  $F(c, -)$ . At the end this process results in a knot  $K'$  such that  $\overrightarrow{u_1u_2}$  has the same direction as  $\overrightarrow{v_1v_2}$  everywhere. Then a linear isotopy from  $K'$  to  $K_0$  as discussed above unknots  $K$  as desired.  $\square$

**Remark 4.14.** Intuitively, the map  $f$  encodes the complexity of how  $K$  is twining around, and the homotopy  $F$  is a process that unwinds it.

## 5. Discrete Morse theory

In this section we set up the foundations of discrete Morse theory, as laid out in , and compare them with their counterparts in section 2.

### 5.1. The analogies

In the classical Morse theory the objects that we study are smooth manifolds. In the discrete setting, these are replaced by **simplicial complexes** ( section 2.1), which can be roughly viewed as combinatorial models of spaces. Particularly, a triangulation of a space makes it into a simplicial complex. By a function on a simplicial complex  $K$ , we mean a function on the set (still denoted as  $K$  by abuse of notation) of the simplexes of  $K$ , i.e. an assignment to each simplex in  $K$  a real number. Throughout, assume  $K$  to be finite.

In section 2 we start with the definition of gradient, but now we no longer have calculus and derivatives at disposal. Instead, we will directly define the notions of gradient field and gradient flow in the following.

**Definition 5.1.** A vector field  $V$  on a simplicial complex  $K$ , is a collection of mutually disjoint pairs  $\{\alpha, \beta\}$  where  $\alpha$  is a facet of  $\beta$  (denoted as  $\alpha < \beta$ ). A flow of  $V$  is a sequence of simplexes

$$\alpha_1, \beta_1, \alpha_2, \beta_2, \dots, \beta_n, \alpha_{n+1} \quad (65)$$

such that  $\forall 1 \leq i \leq n$ ,  $\alpha_i$  and  $\alpha_{i+1}$  are two distinct facets of  $\beta_i$ , and  $\{\alpha_i, \beta_i\}$  is a pair in  $V$ . A flow is said to be closed if  $\alpha_{n+1} = \alpha_1$ . If  $V$  has no closed flow, it is called a **gradient field**.

We intuitively understand a pair as assigning an arrow from  $\alpha$  to  $\beta$ , so that the arrows form the analogy of vector fields and flows in the classical sense. Recall in section 2 that a gradient flow is always associated to a function  $f$ , and along it  $f$  monotonically increases. A gradient flow there is clearly not a loop due to the monotonicity, which motivates the definition above. Moreover, we will soon see that gradient flows come from a certain class of functions on  $K$ .

**Definition 5.2.** A function  $f$  on  $K$  is called a **discrete Morse function** if

1.  $\forall \beta \in K$  a simplex, there is at most one facet  $\alpha$  of  $\beta$  such that  $f(\alpha) \geq f(\beta)$ , i.e.

$$|\{\alpha < \beta | f(\alpha) \geq f(\beta)\}| \leq 1 \quad (66)$$

- for each fixed  $\beta$ .

2.  $\forall \alpha \in K$  a simplex, there is at most one cofacet  $\beta$  of  $\alpha$  such that  $f(\alpha) \geq f(\beta)$ , i.e.

$$|\{\alpha < \beta | f(\alpha) \geq f(\beta)\}| \leq 1 \quad (67)$$

- for each fixed  $\alpha$ .

This might look unenlightening at the first sight, but we will see that this is exactly the condition so that we can generate a gradient flow out of it.

**Proposition 5.3.** If  $f$  is Morse, then  $\forall \tau \in K$ , at least one of  $\{\alpha < \tau | f(\alpha) \geq f(\tau)\}$  and  $\{\tau < \beta | f(\tau) \geq f(\beta)\}$  is empty.

Proof: Otherwise, suppose there exist a facet  $\alpha$  and a cofacet  $\beta$  of  $\tau$  such that  $f(\alpha) \geq f(\tau) \geq f(\beta)$ . There is a unique  $\tau' \neq \tau$  such that  $\alpha < \tau' < \beta$ . Since  $\{\alpha < \sigma | f(\alpha) \geq f(\sigma)\}$  has at most one element and  $\tau$  is already in it,  $f(\alpha) < f(\tau')$ . Similarly,  $f(\beta) > f(\tau')$ . It follows a contraction that  $f(\alpha) < f(\beta)$ .

**Definition 5.4.** A **critical point** of a Morse function  $f$  is a simplex  $\tau \in K$  such that both  $\{\alpha < \tau | f(\alpha) \geq f(\tau)\}$  and  $\{\tau < \beta | f(\tau) \geq f(\beta)\}$  are empty.

**Example.** The dimension function sending  $\tau$  to  $\dim \tau$  is a Morse function, all simplexes are critical points.

We associate a vector field  $V(f)$  to a Morse function  $f$  involving all the non critical points:  $\forall \tau \in K$  non critical, we have either a unique facet  $\alpha$  of  $\tau$  s.t.  $f(\alpha) \geq f(\tau)$  or a unique cofacet  $\beta$  of  $\tau$  s.t.  $f(\tau) \geq f(\beta)$ , we then pick the pair  $\{\alpha, \tau\}$  or  $\{\tau, \beta\}$ . Note that the  $\alpha$  or  $\beta$  will also be non critical, so the pairs thus picked form a mutually disjoint collection.

**Proposition 5.5.**  $V(f)$  is a gradient field. Conversely, every gradient field  $V$  coincides with  $V(f)$  for some Morse function  $f$ .

Proof: For the first statement, suppose that

$$\alpha_1, \beta_1, \alpha_2, \beta_2, \dots, \beta_n, \alpha_{n+1} = \alpha_1 \quad (68)$$

is a closed flow of  $V(f)$ . Since  $\{\alpha_i, \beta_i\}$  is a pair,  $f(\alpha_i) \geq f(\beta_i)$ . Since  $\{\alpha < \beta_i | f(\alpha) \geq f(\beta_i)\}$  has at most one element and  $\alpha_i$  is already in it,  $f(\alpha_{i+1}) < f(\beta_i)$ . It follows a contradiction that  $f(\alpha_1) \geq f(\beta_1) > f(\alpha_2) \geq \dots > f(\alpha_{n+1}) = f(\alpha_1)$ .

For the second statement, define a graph  $\mathcal{K}$  with  $K$  being the vertex set, and two vertices  $\alpha$  and  $\beta$  are joint by an edge if  $\alpha < \beta$ . Make  $\mathcal{K}$  into a directed graph as follows: let  $\alpha \rightarrow \beta$  if  $\{\alpha, \beta\}$  is a pair in  $V$ , and  $\beta \rightarrow \alpha$  if else. We claim that  $\mathcal{K}$  is loopless, so that the topological sorting algorithm ([13]) gives a function  $f$  on (the vertices of)  $\mathcal{K}$  decreasing along each edge, hence  $f$  is a Morse function on  $K$  with  $V(f) = V$ . To prove the claim, notice that along an edge  $e$ , the dimension of the simplexes increases by 1 if  $e$  is a pair in  $V$ , and decreases by 1 if not. It follows that a loop (if exists) of length  $n$  must have exactly  $\frac{n}{2}$  edges in  $V$  and  $\frac{n}{2}$  edges not in  $V$ . But there can not be adjacent edges in  $V$ , since the pairs in  $V$  are mutually disjoint. So in the loop the edges in  $V$  and not in  $V$  appear alternatively. This is exactly a closed flow, contradicting to the assumption that  $V$  is a gradient field. We conclude.

**Remark 5.6.** The analogy in the definition of  $V(f)$  is that "critical points are where the gradient field vanishes". The Morse lemma in section 2 has no analogue here. As for the index, we can just take it to be the dimension of the corresponding simplex. Recall that "index of critical point =

dimension of cell". This slogan is justified by the next result, which is an analogue of the combination of the two fundamental theorems of Morse theory.

## 5.2. Homology from discrete Morse function

**Theorem 5.7.** Suppose that a Morse function  $f$  on  $K$  has  $k$  critical points of dimensions  $\lambda_1 \leq \dots \leq \lambda_k$  respectively, then  $K$  has the homotopy type of a cell complex  $e^{\lambda_1} \cup \dots \cup e^{\lambda_k}$ .

Proof: We are going to use a simple fact: a simplex deformation retracts onto its horn. Thus if  $\alpha$  is a free facet of  $\beta$  (meaning  $\alpha$  has no cofacet expect  $\beta$ ), then  $K$  deformation retracts onto  $K - \alpha - \beta$ .

We prove by induction on  $|K|$ . If  $|K| = 1$ , then  $K$  is a singleton, the result is trivial. For the induction step, among all the top dimensional simplexes, let  $\tau$  be one with maximal  $f$  value. There are two cases:

1.  $\tau$  is critical. Then  $K - \tau$  is again a simplicial complex, and the critical points of  $f|_{K-\tau}$  are exactly those (except  $\tau$ ) of  $f$ . By the inductive assumption,  $K - \tau \simeq e^{\lambda_1} \cup \dots \cup e^{\lambda_{k-1}}$ . Since  $\tau$  is a  $\lambda_k$ -cell attached to  $K - \tau$  along the boundary, by the homotopy extension property ([10] proposition 0.16),  $K \simeq e^{\lambda_1} \cup \dots \cup e^{\lambda_k}$ .
  2.  $\tau$  is non critical. Then let  $\sigma$  be the facet of  $\tau$  such that  $f(\sigma) \geq f(\tau)$ . We claim that  $\sigma$  is a free facet of  $\tau$ , so that  $K - \tau - \sigma$  is again a simplicial complex, and the critical points of  $f|_{K-\tau-\sigma}$  are exactly those of  $f$ . Hence the result follows from the inductive assumption and the fact that  $K - \tau - \sigma \simeq K$ .
- To prove the claim, suppose  $\tau'$  is another cofacet of  $\sigma$ , then  $\tau'$  is top dimensional, so  $f(\tau') \leq f(\tau)$  by the assumption on  $\tau$ . But then  $\{\sigma < \gamma | f(\sigma) \geq f(\gamma)\}$  already has two elements  $\tau$  and  $\tau'$ , a contradiction!

Not only we know  $K$  is homotopy equivalent to  $e^{\lambda_1} \cup \dots \cup e^{\lambda_k}$ , we can also read off the gluing information of the latter from  $f$ . In fact, the corresponding cellular complex can be recovered from analyzing the gradient dynamics of  $f$ , and the homology groups  $H_*(K)$  of (the underlying space of)  $K$  can be computed Morse theoretically. Let  $S_p$  denote the set of the dimension- $p$  critical points (i.e. dimension- $p$  cells in the cell complex) and  $C_p$  denote the free abelian group generated by  $S_p$ . For each critical point, we fix an orientation of the simplex. The boundary maps  $\partial: C_* \rightarrow C_{*-1}$  are determined by the formula

$$\partial(x) = \sum_{y \in S_{p-1}} n_{xy} y, \quad x \in S_p \quad (69)$$

where  $n_{xy} \in \mathbf{Z}$  is a "signed counting" of the gradient flows from  $x$  to  $y$  defined as follows. By a gradient flow  $\gamma$  "from  $x$  to  $y$ ", we mean a gradient flow

$$\alpha_1, \beta_1, \alpha_2, \beta_2, \dots, \beta_n, \alpha_{n+1} \quad (70)$$

of  $V(f)$  such that  $\alpha_1 < x$  and  $\alpha_{n+1} = y$ . Since the orientations of a simplex and of a facet determine each other in a canonical way, we can "push the orientation" of  $x$  along  $\gamma$  to  $y$ . If the resulting orientation at the end agrees with the preexisting orientation of  $y$ , we set the indicator  $m(\gamma) = 1$ , otherwise we set  $m(\gamma) = -1$ . Finally,  $n_{xy}$  is the sum of  $m(\gamma)$  as  $\gamma$  ranges over all gradient flows from  $x$  to  $y$ .

**Theorem 5.8.** ([12] theorem 7.3.) The boundary maps so defined make

$$\dots \xrightarrow{\partial} C_{p+1} \xrightarrow{\partial} C_p \xrightarrow{\partial} C_{p-1} \xrightarrow{\partial} \dots \quad (71)$$

into a chain complex, whose homology computes the homology of  $K$ .

**Remark 5.9.** We will not present a proof (see [14] chapter 8), instead we use a concrete example to illustrate the theory. This result is also an analogue of a concept in classical Morse theory known as the Morse-Smale-Witten complex ([7] section 3), though not mentioned in section 2. The homology is an important topological invariant of simplicial complexes, and has many applications to the real world. For example, in topological data analysis one can compute the persistent homology using discrete Morse theory ([14] chapter 5).

**Example.** Consider below a triangulation of the torus surface.

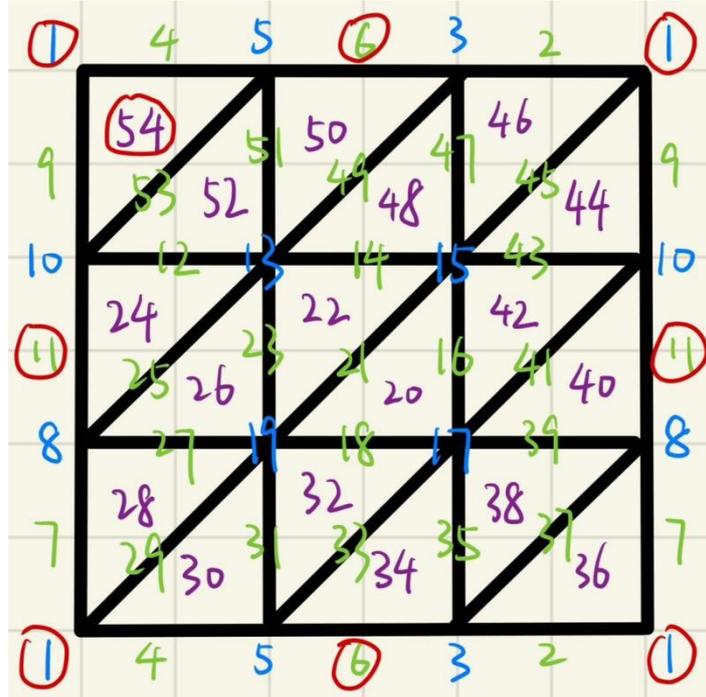


Figure 2: The torus  $T^2$ , with the opposite edges of the largest square identified.

This simplicial complex has 9 vertices, 27 edges and 18 faces, labelled by integers from 1 to 54. This is viewed as a function  $f$  with image  $\{1, 2, \dots, 54\}$ , which can be verified to be a discrete Morse function with 4 critical points encircled in the figure. We describe the gradient field  $V(f)$ : it consists of 25 pairs labelled by  $(i, i + 1)$  that do not contain 1, 6, 11, 54. In fact, tracing the numbers in reverse, this is exactly a process that collapses the 25 pairs in order along gradient flow, so that the remaining is a cell complex with one 0-cell, two 1-cells and one 2-cell. As shown by the fundamental theorem, this recovers a cell decomposition of  $T^2$ . To obtain the homology, we count the gradient flows:

1. From 6 to 1, there are two flows  $5 \rightarrow 4 \rightarrow 1$  and  $3 \rightarrow 2 \rightarrow 1$  that cancel out in orientation.
2. From 11 to 1, there are two flows  $10 \rightarrow 9 \rightarrow 1$  and  $8 \rightarrow 7 \rightarrow 1$  that cancel out in orientation.
3. From 54 to 6, there are two flows  $53 \rightarrow 52 \rightarrow 51 \rightarrow 50 \rightarrow 6$  and  $53 \rightarrow 52 \rightarrow \dots \rightarrow 39 \rightarrow 38 \rightarrow 35 \rightarrow 34 \rightarrow 6$  that cancel out in orientation.
4. From 54 to 11, there are two flows  $53 \rightarrow 52 \rightarrow \dots \rightarrow 40 \rightarrow 11$  and  $53 \rightarrow \dots \rightarrow 38 \rightarrow 35 \rightarrow \dots \rightarrow 24 \rightarrow 11$  that cancel out in orientation.

It follows that all the boundary maps in the chain complex are zero, hence  $H_*(T^2) = C_*$  and we have the homology groups of the torus:

$$H_2(T^2) \cong \mathbf{Z}, H_1(T^2) \cong \mathbf{Z} \oplus \mathbf{Z}, H_0(T^2) \cong \mathbf{Z}. \quad (72)$$

## 6. Conclusion

we gave a survey of both classical Morse theory and discrete Morse theory, and presented their applications. For the classical Morse theory, it can be applied to prove a theorem on the total curvature of knots. For the discrete Morse theory, we applied it to compute the simplicial homology of torus. We outlined the analogies between the two theories entitled "Morse theory", and examined the correspondence of the notions thereof.

## References

- [1] István Fáry. *Sur la courbure totale d'une courbe gauche faisant un noeud. Bulletin de la société mathématique de France*, 77:128–138, 1949.
- [2] John W Milnor. *On the total curvature of knots. Annals of Mathematics*, pages 248–257, 1950.
- [3] Anton Petrunin and Stephan Stadler. *Six proofs of the fáry-milnor theorem. arXiv preprint arXiv:2203.15137v2*, 2023.
- [4] WB Raymond Lickorish. *An introduction to knot theory, volume 175. Springer Science & Business Media*, 1997.
- [5] Elsa Abbena, Simon Salamon, and Alfred Gray. *Modern differential geometry of curves and surfaces with Mathematica. Chapman and Hall/CRC*, 2017.
- [6] John W Milnor. *Morse theory. Number 51. Princeton university press*, 1963.
- [7] Michele Audin, Mihai Damian, and Reinie Ern e. *Morse theory and Floer homology, volume 2. Springer*, 2014.
- [8] John M Lee. *Smooth manifolds. Springer*, 2012.
- [9] Gerald Teschl. *Ordinary differential equations and dynamical systems, volume 140. American Mathematical Soc.*, 2012.
- [10] Allen Hatcher. *Algebraic Topology. Cambridge University Press*, 2002.
- [11] Lawrence Craig Evans. *Measure theory and fine properties of functions. Routledge*, 2018.
- [12] Robin Forman. *A user's guide to discrete morse theory. S eminaire Lotharingien de Combinatoire [electronic only]*, 48:B48c–35, 2002.
- [13] Arthur B Kahn. *Topological sorting of large networks. Communications of the ACM*, 5(11):558–562, 1962.
- [14] Nicholas A Scoville. *Discrete Morse Theory, volume 90. American Mathematical Soc.*, 2019.