A lower bound on critical points of the electric potential of a knot

Max Lipton
Cornell University
Malott Hall, 301 Tower Road
Ithaca, NY 14850
ml2437@cornell.edu

ABSTRACT

Consider a knot $K$ in $S^3$ with charge uniformly distributed on it. From the standpoint of both physics and knot theory, it is natural to try to understand the critical points of the potential and their behavior.

We show the number of critical points of the potential is at least $2t(K) + 2$, where $t(K)$ is the tunneling number, defined as the smallest number of arcs one must add to $K$ such that its complement is a handlebody. The result is proven using Morse theory and stable manifold theory.

Keywords: electrostatics, physical knots, Morse theory

Mathematics Subject Classification 2010: 57M25, 57M27

1. Introduction

Our problem of interest is to analyze the zeros of the electric field around a charged, knotted wire. Let $K \subseteq \mathbb{R}^3 \subseteq S^3$ be a smooth knot parametrized by the curve $r(t)$, $t \in [0, 2\pi]$. We will take the convention that $S^3$ is the union of $\mathbb{R}^3$ and a single compactifying point at infinity. With a choice of units, the electric potential between a point $k \in K$ and a point $x$ at a distance $R$ from $k$ is proportional to $R^{-1}$. It therefore makes sense to define the electric potential $\Phi : S^3 - K \to \mathbb{R}$, on the complement of $K$ by the line integral

$$\Phi(x) = \int_{k \in K} \frac{dk}{|x - k|} = \int_0^{2\pi} \frac{|r'(t)|}{|x - r(t)|} dt, \ x \in \mathbb{R}^3.$$  \hspace{1cm} (1.1)

As usual, we set $\Phi(\infty) = 0$. By differentiating in the integral sign with respect to $x$, we can see the electric potential is smooth and harmonic. The electric field is defined by $E = -\nabla \Phi$. We want to describe the critical points of the potential (equivalently, the zeros of the electric field) and their behaviors. Some conventions use the negative of the potential, so that the electric field points towards the knot, but it is more convenient for our purposes to work with a nonnegative potential function.
Our problem differs from much of the existing research on physical knots. Many works, such as Cantarella et al. [3], focus entirely on mechanical questions like the rope length needed to tie a knot. Even within the subdiscipline of electrostatic knot theory, physicists have debated on how to define the electric potential of a knot. Some papers consider knots with thickness, where the potential is defined as a volume integral, like in Buck et al. [2]. The definition given here could be criticized for the fact that it does not blow up when the knot self-intersects [8]. The blow up property may be desirable because it acts as an obstruction to varying the knot type, but we elect to use the given potential because of its harmonicity and smoothness. These facts are crucial to our proofs. Simon [8] has a thorough account of the historical debate.

We obtain a lower bound for the number of critical points of the electric potential based on a well known topological invariant called the tunneling number. The tunneling number was originally introduced by Clark [4], and remains an active topic of knot theory research [1]. Proving our result uses Morse theory and stable manifold theory.

2. Preliminary Definitions and Lemmas

The theorems of Morse theory require us to work on a compact manifold, so in the sequel we will define the knot complement of $K$ in $S^3$ to be the complement of an open tubular neighborhood of $K$ of sufficiently small radius. The theorems from Morse theory we invoke will still hold on this compact manifold with boundary because the potential function is proper, with gradient transversely intersecting the boundary. We will still denote our domain by $S^3 - K$.

A critical point of a smooth real valued function on a manifold is a point $p$ such that the differential $df(p)$ is zero. The critical set of $f$ is the set of critical points, and is denoted $\text{Crit}(f)$. We say $f$ is Morse if its critical points are nondegenerate, which means the Hessian matrix $H(f)$ of second partial derivatives is nonsingular. The index of $p \in \text{Crit}(f)$ is the number of negative eigenvalues of $H(f)$, which is invariant under the choice of local coordinates. We denote the set of critical points of index $i$ by $\text{Crit}_i(f)$. If the Morse function is fixed, we denote the index of $p$ by $\lambda(p)$ and we denote the number of critical points of index $i$ by $m_i$.

We write $W^S(p)$ and $W^U(p)$ to denote the stable and unstable manifolds of $p \in \text{Crit}(f)$ respectively. Recall that the stable manifold (resp. unstable manifold) of $p$ is the set of all points which flow to $p$ along the gradient as time tends to infinity (resp. negative infinity). If $f$ is Morse, the dimension of $W^S(p)$ is $\lambda(p)$ and the dimension of $W^U(p)$ is $\dim M - \lambda(p)$.

The Morse Reconstruction Theorem states the domain of a Morse function on a compact manifold can be expressed as a cell complex by attaching closed discs with dimensions given by the indices of the critical points. The attaching maps are obtained by a process known as surgery, but we will not need to discuss the attaching maps in any further detail.
In the space of all smooth real valued maps, under a suitable function space topology, the set of Morse functions is dense. Morse functions are structurally stable in the sense that sufficiently small perturbations do not affect the size of the critical set or the indices. Therefore, we may assume $\Phi$ is Morse by perturbing it slightly should it not be Morse, and the way in which we perturb will not affect the indices of the critical points.

The Morse inequalities state for a fixed Morse function, $m_i \geq b_i$, where $b_i = \dim H_i(M)$ is the $i^{th}$ Betti number of our domain manifold. We will need the stronger result which states

$$
\dim M \sum_{i=0}^{\dim M} (-1)^i m_i = \dim M \sum_{i=0}^{\dim M} (-1)^i b_i = \chi(M),
$$

with $\chi(M)$ denoting the Euler characteristic of $M$. Proofs of all of the above results can be found in Nicolaescu [7].

Before stating the main theorem, we need to prove a few preliminary lemmas.

**Lemma 2.1.** For all knots $K$, $H_i(S^3 - K) = \mathbb{Z}$, for $i = 0, 1$ and $H_i(S^3 - K) = 0$ for $i \geq 2$.

**Proof.**

Obviously, $H_i(S^3 - K) = 0$ for $i \geq 4$ because $S^3 - K$ is a 3-manifold. The remainder of the proof is a straightforward application of the Mayer-Vietoris sequence. Let $A$ be a tubular neighborhood of $K$, which is homeomorphic to a solid torus, and therefore has the homotopy type of a circle. Let $B = S^3 - K$, whose homology we wish to compute. Then $A \cup B = S^3$, and $A \cap B$ is homeomorphic to a torus $T^2$ embedded in $\mathbb{R}^3$.

Consider the following portion of the Mayer-Vietoris sequence:

$$
\cdots \to H_3(T^2) \to H_3(A) \oplus H_3(B) \to H_3(S^3) \oplus H_2(T^2) \to \cdots
$$

By definition, the map $\partial_*$ is induced by taking a 3-chain in $S^3$ and intersecting with $T^2$, so notice $\partial_*$ maps the fundamental class of $S^3$ to that of $T^2$. Therefore, $\partial_*$ is an isomorphism. Thus, by exactness, $H_3(S^3 - K) = 0$.

Consider this segment of the Mayer-Vietoris sequence:
\[ H_2(S^3) \xrightarrow{\partial_*} H_2(T^2) \xrightarrow{\phi_*} H_2(A) \oplus H_2(B) \xrightarrow{\psi_*} H_2(S^3) \xrightarrow{\partial_*} H_2(S^3) \xrightarrow{\partial_*} \cdots \]

Since \( \partial_* \) is an isomorphism, we have that \( \phi_* \) is the zero map by exactness. As \( \psi_* \) is also a zero map, it follows that \( H_2(S^3 - K) \) must also be zero for exactness to hold. Finally, consider this last segment of the Mayer-Vietoris sequence:

\[ \cdots \rightarrow H_2(S^3) \rightarrow H_1(T^2) \rightarrow H_1(A) \oplus H_1(B) \rightarrow H_1(S^3 - K) \rightarrow 0 \]

As the middle two terms are surrounded by zeros, it follows they are isomorphic. This implies \( H_1(S^3 - K) \) must be \( \mathbb{Z} \) in order for the direct sum to have \( \mathbb{Z} \)-rank 2. Finally, we note that as \( S^3 - K \) is connected, \( H_0(S^3 - K) = \mathbb{Z} \).

From the previous computation, along with (2.1), we can deduce the following lemma.

**Lemma 2.2.** For all knots \( K \), and with \( \Phi \) defined above, \( m_1 - m_2 = 1 \).

**Proof.** Equation (2.1) states the Euler characteristic equals the alternating sum of the \( m_i \)'s. In other words, \( \chi(S^3 - K) = \sum_{i=0}^{3} (-1)^i m_i \). From the homology computation, we can see that \( \chi(S^3 - K) = 1 - 1 = 0 \). Since \( \Phi \) is harmonic, every critical point has index 1 or 2, save for the point at infinity, which has index 0. We can conclude \( m_0 - m_1 + m_2 - m_3 = 0 \), or equivalently, \( m_1 - m_2 = 1 \) as desired.

A handlebody is a topological space homotopic to the three dimensional ball with solid handles attached (by “attaching handles” we mean there are copies of \( D^2 \times [0,1] \) where the boundary discs \( D^2 \times 0 \) and \( D^2 \times 1 \) are embedded on the boundary of the three-ball). Given a knot \( K \subseteq S^3 \), the tunneling number \( t(K) \) is the least number of arcs we must add to \( K \) such that the complement in \( S^3 \) is a handlebody. A collection of arcs with this property is known as a tunneling.

In the proof of our result, we shall use the above definition of the tunneling number, but there is an equivalent definition that is more visually intuitive. A Heegaard splitting of a three-manifold \( M \) is an embedding of a closed, compact, and orientable surface \( H \) such that the interior and exterior of \( H \) in \( M \) are both handlebodies. We say the genus of \( H \) is the genus of the splitting. Below, we state
two results from low dimensional topology. For proofs and exposition, see Johnson [5].

**Theorem 2.3.** Let $H$ be an unknotted embedding of a genus $g$ surface in $S^3$. That is, let $H$ be the boundary of a tubular neighborhood around a wedge sum of $g$ unknotted circles. Then $H$ defines a Heegaard splitting.

**Theorem 2.4.** Genus $g$ Heegaard splittings of $S^3$ are unique up to isotopy.

Theorem 2.4 is known as Waldhausen’s Theorem. Clearly, a tunneling of a knot defines a Heegaard splitting. Therefore, we can view the tunneling number of $K$ as the least number of arcs we must add to $K$ such that it is isotopic to a wedge sum of unknotted circles. We can immediately deduce that tunnelings and the tunneling number always exist.

**Lemma 2.5.** Every knot $K$ has a tunneling number.

**Proof.** Take a diagram of $K$ with finitely many crossings. Over each crossing, introduce an arc connecting the top and bottom strands. Collapse each arc so that the top and bottom strands intersect, and project onto the diagram’s plane so that we are left with a wedge sum of say, $g$ circles. By Theorem 2.3, it follows that the complement in $S^3$ is also a handlebody with $g$ handles.

**Remark 2.6.** We just proved the tunneling number is bounded above by the crossing number, the least number of crossings needed in a knot diagram of $K$.

As some elementary examples, the tunneling number of the unknot is zero, and the tunneling number of the trefoil is one. See Clark [4]. Indeed, the tunneling number need not be the crossing number. We now come to the main result of this article.

**Theorem 2.7.** The electric potential of $K$ has at least $2t(K) + 2$ critical points, counting the point at infinity.

3. **Proof of Theorem 2.7**

To prove the result, we construct a tunneling with $m_2$ arcs. This proves $m_2 \geq t(K)$. Applying Lemma 2.2, we get that $m_0 + m_1 + m_2 + m_3 \geq 1 + (t(K) + 1) + t(K) + 0 = 2t(K) + 2$ as desired.

3.1. **Construction of the tunneling**

We will apply the Morse Rearrangement Lemma to allow us to make some assumptions about $\Phi$. A proof can be found in Nicolaescu [7]. The theorem states we can find a smooth $\hat{\Phi} : S^3 - K \to \mathbb{R}$ satisfying the following properties:
• Φ and ˆΦ share the same critical points, and each critical point has the same index.
• Suppose p and q are distinct critical points. Then ˆΦ(p) ̸= ˆΦ(q). If ˆΦ(p) < ˆΦ(q), then λ(p) ≤ λ(q).
• Inside of a neighborhood of each critical point, the gradient flows for Φ and ˆΦ are identical.
• If γ is an integral curve of ∇Φ, then Φ(γ(t)) is strictly increasing in t. A vector field with this property is called gradient-like with respect to Φ.
• For p, q ∈ Crit(ˆΦ), WS(p) and Wu(q) intersect transversely. Morse functions with this property are called Morse-Smale.

This theorem allows us to perturb the values of the critical points so we can reorder them ascending by index without affecting the topological data encoded by the original potential. At this point, we are not necessarily working with the physical potential whose formula is given in (1.1), but for simplicity we will still refer to the perturbation as Φ.

Our rearrangement restricts the limiting behavior of trajectories. Let γ : (−∞, ∞) → S3 − K be a trajectory for Φ. When t → −∞, Φ(γ(t)) strictly decreases, but it is bounded below by 0, by assumption. Should lim t→−∞ Φ(γ(t)) = 0, then lim t→−∞ γ(t) = ∞, the point of infinity on S3, because its the only point in the knot complement with zero potential. Otherwise, lim t→−∞ γ(t) is a critical point.

Similarly, we can deduce that for t → ∞, either γ(t) tends to a critical point or K, since Φ(γ(t)) is strictly increasing.

Should both ends of γ be critical points, then we know the index of the critical point at t = −∞ is less than or equal to that of the critical point at t = ∞.

For p^2_1, . . . , p^2_m_2 ∈ Crit2(Φ), let Γ_i = W^U(p^2_i). Notice each Γ_i is a union of two trajectories leaving p^2_i, since the unstable manifold has dimension 1. Since Φ strictly increases along trajectories, we have that the trajectories will either tend to K or to another critical point of index 2 as t → ∞. However, should either end of Γ_i tend to a critical point q ∈ Crit2(Φ), then the corresponding trajectory will be a submanifold WS(q). However, WS(q) and Γ_i are 2 and 1-submanifolds respectively, and for them to intersect transversely as per our assumption, their intersection cannot be more than 0 dimensions. Therefore, we conclude both ends of Γ_i eventually reach the tubular neighborhood of the knot. Our tunneling is only concerned with the arc outside of the tubular neighborhood, so we can assume Γ_i is defined only on a compact interval.

For p^1_1, . . . , p^1_m_1 ∈ Crit1(Φ), let Θ_j = WS(p^1_j). Analogous to before, each Θ_j is a union of two trajectories tending to p^1_j. Similar reasoning will tell us that both ends tend to the point at infinity. Indeed, each Θ_j is a union of two trajectories tending towards a critical point p^1_j of index 1. As t → −∞, Φ will strictly decrease along these trajectories, so we know that the negative infinite limits of these trajectories must either be another critical point of index 1, or the point at infinity. By the
Fig. 1. The arcs we add to $K$ are the unstable manifolds associated to critical points of index 2. Note that this diagram does not necessarily depict the specific situation accurately for the trefoil.

transversality assumption of the stable and unstable manifolds, we cannot have that the endpoints of $\Theta_j$ are critical points of index 1. Therefore, both ends are at the point at infinity. We may view the union of all the $\Theta_j$ arcs as a wedge sum of circles at the point of infinity, which we denote $\bigvee_{j=1}^{m_1} \Theta_j$.

Notice $\bigvee_{j=1}^{m_1} \Theta_j$ is homotopy equivalent to a handlebody. Using a standard maneuver from differential topology, we will flow along the (negative) gradient to perform a deformation retraction from $S^3 - (K \cup \Gamma_1 \cup \cdots \cup \Gamma_{m_2})$ to $\bigvee_{j=1}^{m_1} \Theta_j$. We will opt to work with smooth tubular neighborhoods of both of these spaces, but there is a crucial technical lemma we must prove before proceeding.

3.2. Constructing a smooth boundary around the tunneling

In this subsection, we prove the following lemma.

Lemma 3.1. There are tubular neighborhoods of both $K \cup \Gamma_1 \cup \cdots \cup \Gamma_{m_2}$ and $\bigvee_{j=1}^{m_1} \Theta_j$ with smooth boundary such that the gradient vector field points inwards and outwards respectively.

Proof.

It is a standard fact from electrostatics that neighborhoods of $K$ and the point at infinity exist such that the gradient points inwards and outwards respectively. This proof will make use of the tubular flow lemma, which states for every regular point of $S^3 - K$, there are local coordinates such that the gradient flow takes the form constant form $\nabla_{(x_0,y_0,z_0)} \Phi = (1,0,0) = \frac{\partial}{\partial x}$. Take local coordinates centered at a regular point of $\Gamma_i$ so that the portion of the arc inside our coordinate chart is mapped to the path of unit speed along the $x$-axis, $\gamma(t) = (t,0,0)$. 


Fig. 2. A sketch depicting the tubular neighborhood around a regular point of $\Gamma_i$ in specially chosen local coordinates which makes the gradient parallel. This is a projection to the $xz$-plane.

Around a segment of the $x$-axis, we can choose our tubular neighborhood of $\Gamma_i$ to be the interior of the slanted tube defined by $y^2 + z^2 = (\frac{1}{2}x + 1)^2$, where we possibly restrict our local coordinates so $-1 < x < 1$. See Fig. 2. Notice the gradient points inwards from the boundary. If we expand around a regular point on a $\Theta_j$ arc, we can reflect the tube in the $x$ direction so the gradient points outwards. Note that the properties of a vector field pointing inwards and outwards from a boundary are invariant under a change of coordinate charts in an orientation-preserving atlas.

Now suppose we want to take local coordinates around a critical point $p^2_i$ of $\Gamma_i$, which has index 2. By Lemma 2.2 of [6], we can find local coordinates centered at the critical point such that $\Gamma_i$ corresponds to the $x$-axis, and $\Phi$ takes the form $\Phi(x_0, y_0, z_0) = \frac{1}{2}(x_0^2 - y_0^2 - z_0^2) + c$, for some constant $c$. Thus, $\nabla \Phi$ takes the form $\nabla \Phi(x_0, y_0, z_0) = (x_0, -y_0, -z_0)$. Consider the tube around the $x$-axis defined by $y^2 + z^2 = 1$. See Fig. 3. Using an abuse of notation, we will refer to this tube as $\partial \Gamma_i$. We will show the gradient points inside the tube, as Figure 3 shows. At a point $(x_0, y_0, z_0) \in \partial \Gamma_i$, the tangent plane is spanned by $(1, 0, 0)$ and $(0, -z_0, y_0)$. As the point varies on the tube, this oriented basis of the tangent plane varies smoothly, thus defining an orientation of the tube. The triple of tangent vectors $\{(1, 0, 0), (0, -z_0, y_0), \nabla \Phi(x_0, y_0, z_0)\}$ therefore defines a smooth choice of orientation for the three dimensional ambient tangent spaces surrounding the tube. Therefore, if the gradient points inside the tube at one point of the tube, it does so throughout the whole tube. For example, at the point $p = (0, 1, 0) \in \partial \Gamma_i$,
Fig. 3. A sketch depicting the tubular neighborhood around a critical point on $\Gamma_i$, in specially chosen local coordinates such that $\Phi$ is a quadratic with signature $(1, 2)$. This is a projection to the $xz$-plane.

$\nabla \Phi(p) = (0, -1, 0)$ clearly points inside the tube. The existence of a surface surrounding the arc with which the gradient points inwards is a significant topological obstruction to proving our main theorem. The construction would not be possible if the critical point had index 1, or if we were asked to place a tube around another axis. The case for a critical point of $\Theta_i$ is analogous, and we get the result that the gradient points outwards.

By compactness of the arcs $\Gamma_i$ and $\Theta_j$ we only need to construct finitely many tubes around arc segments to encapsulate the entire arc. When the tubes cover the same part of the arc, we may have to shrink the radius of one of the tubes so the boundaries will intersect, but the resulting tube will still have the gradient pointing in the correct direction. By taking the final boundary to be the points of minimal radial distance to the arc, we obtain a connected boundary to the tubular neighborhood that is only piecewise smooth. Likewise, the tubular neighborhoods around the $\Gamma_i$ arcs intersect the tube around $K$, and the gradient points inwards on the boundary of the union. This construction is sufficient to prove the lemma. To get a smooth boundary out of the piecewise smooth boundary, one could either make a density argument in a space of manifolds [7], justify why the theorems work in the piecewise smooth case [6], or use mollifiers to smooth out the kinks [9]. For
the sake of space, we omit the technical details.

For reasons that will be clear when we perform the deformation retraction, we will want to include a ball around the point at infinity in the tubular neighborhood. In the standard $\mathbb{R}^3$ coordinates, this is the complement of a large open ball. We can assume this neighborhood around $\infty$ contains $\Phi^{-1}([0,\varepsilon])$ for some $\varepsilon > 0$. On the boundary sphere, we can again use mollifiers to connect the tube smoothly whilst preserving their orientations against the gradient flow.

3.3. The deformation retraction to a handlebody

The final step is to use the flow of $E$ to perform the deformation retraction. We will use the closed tubular neighborhoods described in Lemma 3.1 as an alternative to just the knot with arcs attached and the wedge of circles. Let $A$ be the aforementioned tubular neighborhood of the knot with the $\Gamma_i$ arcs attached, and let $B$ denote the aforementioned tubular neighborhood of the $\Theta_j$ arcs connected to a ball around the point at infinity.

Let $E(x,t)$ be the flow of the negative gradient. The point $E(x,t)$ refers to the location of the path at time $t$ starting from the unique integral curve starting at $x$. First, we prove every point in $S^3 - A$ will eventually flow to $B$. Suppose $x \in S^3 - A$. There are three possibilities for the limit of the negative gradient flow of $x$: it could flow to $K$, it could flow to a critical point of index 1 or 2, or it could flow to the point at infinity. Since the negative gradient points outwards from the boundary of $A$, $x$ cannot flow to $K$ or a critical point of index 2. Therefore, $x$ either flows to a critical point of index 1, or to the point at infinity, which means that $x$ eventually flows to $B$. Furthermore, since the negative gradient points into $B$, once $x$ enters $B$, it will never leave.

By smoothness of the boundary of $B$, the function which assigns each $x \in S^3 - A$ the minimum time $t$ such that $E(x,t) \in B$ is smooth. Call this function $C(x)$. Notice $C$ assigns 0 to each point already in $B$. By compactness of the domain, the function reaches a maximum value $C_{\text{max}}$. Define a homotopy $H$ on $(S^3 - A) \times [0,\infty)$ by $H(x,t) = E(x, \min(t, C(x)))$, which we can see is continuous (in fact, it is smooth). Also notice that for $x \in B$, $H(x,t) = x$ for all $t$. Running this homotopy on the time interval $[0, C_{\text{max}}]$ completes the deformation retraction. This completes the proof of our main theorem. □

Acknowledgments

The author would like to thank his advisor, Steven Strogatz, for introducing him to this problem, and for helping motivate the connections to physics. He would also like to thank Greg Buck, Rennie Mirollo, Tim Riley, and Alex Townsend for their helpful discussions, comments, and suggestions. Finally, he would like to thank the participants of the RTG Math Communications Seminar (of whom there are too many to list) for their feedback on the writing.
This work was funded in part by the NSF RTG Grant entitled Dynamics, Probability, and Partial Differential Equations in Pure and Applied Mathematics, DMS-1645643.

References