ADDITIVITY OF BRIDGE NUMBERS OF KNOTS

JENNIFER SCHULTENS

ABSTRACT. We provide a new proof of the following results of H. Schubert: If K is a satellite knot with companion J and pattern $(\hat{V}, L)$ with index k, then the bridge numbers satisfy the following equation: $b(K) \geq k \cdot (b(J))$. In addition, if K is a composite knot with summands J and L, then $b(K) = b(J) + b(L) - 1$.

In "Über eine numerische Knoteninvariante" [1], Horst Schubert proved that for a satellite knot K with companion J and pattern of index k, bridge numbers satisfy the inequality $b(K) \geq k \cdot (b(J))$. He also proved that for a composite knot K with summands J and L, the bridge numbers satisfy $b(K) = b(J) + b(L) - 1$. His investigation was motivated by the question as to whether a knot can have only finitely many companions. Together with the fact that the only bridge number one knot is the unknot, his result showed that the answer to this question is yes.

Schubert's result may be recovered by a much shorter proof. Although this proof does not explicitly employ the notion of thin position of a knot, it relies heavily on the idea of rearranging the order in which critical points occur to suit one's purpose, an idea fundamental to the notion of thin position of knots and 3-manifolds. This proof also differs from Schubert's in that it relies on the consideration of Morse functions on S^3 whose level sets are spheres (except for the maximum and minimum) rather than on a foliation of $\mathbf{R}^3$ by planes.

I wish to thank both Ray Lickorish and Marty Scharlemann for independently suggesting that a more modern proof of this theorem would be a welcome addition to the existing literature. I also wish to thank the Department of Pure Mathematics and Mathematical Statistics at the University of Cambridge for its hospitality.

In the following K will always be a knot in S^3 and $h : S^3 \to \mathbf{R}$ a Morse function with exactly two critical points. This last assumption guarantees that h induces a foliation of S^3 by spheres, along with one maximum that we denote by ∞ and one minimum that we denote by $-\infty$.

Definition 1. *If the minima of h_K occur below all maxima of h_K, then we say that K is in bridge position with respect to h. The bridge number of K, $b(K)$, is the minimal number of maxima required for h_K. (Note that this number is independent of whether or not we require K to be in bridge position. Indeed, if h_K has n maxima, then the maxima of h_K can be raised, and the minima of h_K lowered, to obtain a copy of K in bridge position with n maxima.)*

Definition 2. *Let J be a knot in S^3 and denote a small closed regular neighborhood of J by $\tilde{V}$. Let $\hat{V}$ be an unknotted solid torus in S^3 containing a knot L. A map of*

1991 *Mathematics Subject Classification.* 57N10.

Research supported by a grant from the NSF.

$\hat{V}$ into V maps L onto a knot K. We call K a satellite knot, J a companion of K, V the companion torus of K with respect to J and $(\hat{V}, L)$ the pattern (of K with respect to J). The least number of times which a meridian disk of V intersects L is called the index of the pattern.

Definition 3. *Suppose that K is homotopically nontrivially contained in a solid torus V. Set $T = \partial V$. Then V is taut with respect to $b(K)$, if the number of critical points of h_T is minimal subject to the condition that h_K has $b(K)$ maxima.*

Definition 4. *Consider the singular foliation, $\mathcal{F}_T$, of T induced by h_T. Let σ be a leaf corresponding to a saddle singularity. Then σ consists of two circles, s_1, s_2, wedged at a point. If either s_1 or s_2 is inessential in T, then we call σ an inessential saddle. Otherwise, σ is an essential saddle.*

Lemma 1. *(The Pop Over Lemma) Let $h, K, V, \mathcal{F}_T$ be as above. If $\mathcal{F}_T$ contains inessential saddles, then, after an isotopy of T that does not change $b(K)$ or the number of critical points of h_T, there is an inessential saddle σ in $\mathcal{F}_T$ for which the following conditions hold:*

1) s_1 bounds a disk $D_1 \subset T$ such that $\mathcal{F}_T$ restricted to D_1 contains only disjoint circles and one maximum or minimum; and

2) for L the level surface of h containing σ and E the closure of the component of $T - L$ (other than D_1) that contains s_1 (so, for instance, $s_2 \subset \partial E$), D_1 cobounds a 3-ball B with a disk $\tilde{D}_1 \subset L$, such that B does not contain ∞ or $-\infty$, and such that s_2 does not meet B (i.e., such that s_2 lies outside of $\tilde{D}_1$).

Proof. The first condition on σ may be satisfied by choosing σ to be an innermost inessential saddle in $\mathcal{F}_T$. In this case $L - \partial D_1$ consists of two disks, $\hat{D}_1$ and $\hat{D}_2$. Together with D_1, both $\hat{D}_1$ and $\hat{D}_2$ cobound 3-balls $\hat{B}_1, \hat{B}_2$, respectively. One of these 3-balls, say $\hat{B}_2$, contains either ∞ or $-\infty$ and the other contains neither.

If $s_2 \subset \hat{D}_2$, we may take $B = \hat{B}_1$, so suppose $s_2 \subset \hat{D}_1$. Without loss of generality, we may assume that the critical point of D_1 is a maximum. In this case, consider a monotone arc α disjoint from K, beginning at the maximum of D_1, passing only through maxima of T and ending at ∞. Let $a_1, \dots, a_n$ be the points at which α meets T, with a_n the highest such point.

Let β be the subarc of α between a_n and ∞ and let C' be a collar neighborhood of β. After a small isotopy, $T \cap C'$ consists of a small disk $D = a_n \times disk \subset T$. Let C'' be a small 3-ball centered at ∞ that is disjoint from T. Set $C = C' \cup C''$ and consider $T' = (T - D) \cup (\partial C - D)$. This describes an isotopy of T that replaces $\hat{B}_1$ by $\hat{B}_1 \cup C$ and replaces $\hat{B}_2$ by $\hat{B}_2 - C$. After a small tilt which turns $h_{T'}$ into a Morse function, the maximum a_n of h_T has turned into a maximum of $h_{T'}$ at a higher level. No critical points need have been introduced for h_K and the number of critical points of $h_{T'}$ is the same as that of h_T.

By induction, we may assume that α is disjoint from T except at its initial point. Then if $s_2 \subset \hat{D}_1$, this same construction using $\beta = \alpha$ describes an isotopy of T augmenting $\hat{B}_1$ to contain ∞ and shrinking $\hat{B}_2$ to exclude ∞ without introducing any critical points of h_K or h_T. We may then choose B to be the shrunk version of $\hat{B}_2$. □

fig. 1

Lemma 2. *(The Pop Out Lemma) Let $h, K, V, \mathcal{F}_T$ be as above. If V is taut with respect to $b(K)$, then there are no inessential saddles in $\mathcal{F}_T$.*

Proof. Suppose there are inessential saddles. Then let σ be an inessential saddle satisfying the conclusions of Lemma 1. We may assume that D_1 contains a maximum and lies above L. (The other case is analogous.) Here $(K \cup T) \cap int(B)$ can be shrunk and lowered via an isotopy to lie just below $\tilde{D}_1$ (and above any critical points of h_K or h_T below $\tilde{D}_1$). This does not change the nature or number of critical points of h_K or h_T.

Now $D_1 \subset T$ can be replaced by $\tilde{D}_1$ to obtain $\tilde{T}$. After a small tilt, $\tilde{T}$ bounds a solid torus $\tilde{V}$ containing a copy of K with $b(K)$ maxima, and $\tilde{T}$ is isotopic to T, yet $h_{\tilde{T}}$ has two fewer critical points than h_T. (A maximum and an inessential saddle have been cancelled). This contradicts the assumption that V is taut with respect to $b(K)$. ☐

fig. 2

Remark 1. *Consider a bicollar of an essential saddle σ in $\mathcal{F}_T$. It has three boundary components, c_1, c_2, c_3, where c_i is parallel to s_i for $i = 1, 2$. Since $\chi(T) = 0$, it follows that c_3 bounds a disk. If there are no inessential saddles, then the disk bounded by c_3 contains exactly one singular point, a maximum or minimum. We consider this maximum or minimum, m_σ, to be the maximum or minimum corresponding to σ.*

Conversely, if there are no inessential saddles in $\mathcal{F}_T$, then every maximum or minimum corresponds saddle in this way, since $\chi(T) = 0$.

Definition 5. *Suppose σ is a saddle in $\mathcal{F}_T$, r_1 a regular level of h slightly above, and r_2 a regular level of h slightly below $h(\sigma)$. If $|h^{-1}(r_1) \cap V| - |h^{-1}(r_2) \cap V| = 0$, (rather than ± 1), then we call σ a nested saddle.*

Remark 2. *For any saddle σ, $\chi(h^{-1}(r_1) \cap V) - \chi(h^{-1}(r_2) \cap V) = \pm 1$ for r_1, r_2 as above. In particular, for L any level surface above the highest nested saddle, all components of $L \cap V$ are disks.*

Lemma 3. *Let $h, K, V, \mathcal{F}_T$ be as above. If V is taut with respect to $b(K)$, then $\mathcal{F}_T$ has no nested saddles.*

Proof. If there are nested saddles, consider the highest such saddle, σ. By Lemma 2, σ is an essential saddle, so both s_1 and s_2 are essential in T. Set $L_\sigma = h^{-1}(h(\sigma))$. Then $L_\sigma - \sigma$ consists of three disks, A_1, A_2, A_3. Two of these disks, say A_1 and A_2 are exterior to V near σ.

Note that any component of intersection of $A_i \cap V$, for $i = 1, 2$ must be a disk, since there are no nested saddles above L_σ. As V is knotted, and $\partial A_i = s_i$, for $i = 1, 2$, an essential curve, A_i must intersect V in at least one essential disk, for $i = 1, 2$. As they are disks in V bounded by essential curves, they must be meridian disks.

The existence of these meridian disks establishes two facts: 1) The existence of essential saddles in $\mathcal{F}_T$ above σ that are not nested. 2) For any nested saddle $\tilde{\sigma}$, the curves $\tilde{s}_1$ and $\tilde{s}_2$, are essential curves disjoint from these meridian disks of V, and are hence also meridian curves of V.

There must thus be an "adjacent" pair σ_1, σ_2 of essential saddles with σ_1 nested, σ_2 not nested, where "adjacent" means that one component, say C, of $T - (\sigma_1 \cup \sigma_2)$

contains no critical points of h_T. Consider the circles s_1^i, s_2^i whose wedge is σ_i. Without loss of generality, s_1^1 and s_1^2 meet C.

Again without loss of generality, we may assume that σ_1 lies above σ_2 and hence that the component of $T - \sigma_1$ lying above σ_1 and meeting both s_1^1 and s_2^1 is a disk D_3^1. Construct a disk D by adding D_3^1 to C and capping off s_2^1 with a component of $h^{-1}(h(\sigma_1)) - \sigma_1$. (Note that by the discussion above, this latter horizontal portion of D meets K and V.)

We now proceed as in Lemma 1 and Lemma 2. Here $\partial D = s_1^2$, so ∂D divides $h^{-1}(h(\sigma_2)$ into two disks, $\hat{D}_1$ and $\hat{D}_2$, that cobound 3-balls $\hat{B}_1$ and $\hat{B}_2$ with D. As in the proof of Lemma 1, we may assume that $\hat{B}_2$ contains ∞. If $s_2^2 \subset \hat{B}_2$, set $B = \hat{B}_1$. If $s_2^2 \subset \hat{B}_1$, augment $\hat{B}_1$ and shrink $\hat{B}_2$ via an isotopy of T identical to the one used in Lemma 1 and then set $B = \hat{B}_2$ (the shrunk version).

Now shrink $(K \cup T) \cap B$ similarly as in the proof of Lemma 2. However, here $K \cup T$ meets D along its horizontal portion. As $(K \cup T) \cap B$ is shrunk, the horizontal portion of D is lowered while remaining horizontal. In order to accomplish this, the portion of B lying above $h^{-1}(h(\sigma_1))$ must be shrunk horizontally as necessary. In the end, a product neighborhood below the original horizontal portion of D will end up intersecting $K \cup T$ in vertical arcs and surfaces.

As in the proof of Lemma 2, the number of critical points of h_T can be reduced without altering the number of critical points of h_K, contradicting the fact that V is taut with respect to $b(K)$. □

fig. 3

Remark 3. *If V is a knotted solid torus that is taut with respect to $b(K)$ then all saddles are essential and there are no nested saddles. If follows that if $L = h^{-1}(r)$ for some regular value r, then $V \cap L$ consists of disks. In particular, by shrinking these disks to be very small, and by shrinking V so that maxima and minima of T are very close to the essential saddles to which they correspond, V can be considered as a small regular neighborhood of its core.*

Theorem 1. *Suppose K is a satellite knot with companion J, companion torus $\tilde{V}$, pattern $(\hat{V}, L)$ and index k. Then $b(K) \geq k \cdot b(J)$. In addition, if K is the connected sum of two knots K_1 and K_2, then $b(K) = b(K_1) + b(K_2) - 1$.*

Proof. We may assume that V is taut with respect to $b(K)$. Then V is a small regular neighborhood of its core. We obtain a Morse function on (S^3, J) by making V very thin. So $b(J)$ is less than or equal to the number of maxima of $h_{T=\partial V}$.

Consider a maximum of T. It corresponds to a saddle σ, where σ is the wedge of the circles s_1, s_2, bounding level meridian disks $\tilde{D}_1, \tilde{D}_2$ of V. Here $\tilde{D}_1 \cup \tilde{D}_2$ cuts off a 3-ball B_σ of V. For distinct saddles σ_1 and σ_2, B_{σ_1} and B_{σ_2} will be disjoint. Since at least k strands pass through both $\tilde{D}_1$ and $\tilde{D}_2$, there are at least k maxima of K in B_σ. Whence $b(K) \geq k \cdot b(J)$.

In the special case where K is the connected sum of K_1 and K_2, the sattelite construction may be used with K_1 the companion and $(\hat{V}, K_2)$ the pattern (of index 2). Then we still obtain a Morse function on (S^3, K_1) as above. Furthermore, we may construct a Morse function on $(\hat{V}, K_2)$ as follows: First embed $\hat{V}$ unknottedly in S^3 endowed with a Morse function $\tilde{h}$ in such a way that $\partial \hat{V}$ inherits a foliation identical to $\mathcal{F}_T$ and such that $\tilde{h}_{K_2}$ is identical to $h_{K_1 \# K_2}$.

If $\tilde{h}_{\partial\hat{V}}$ has more than one maximum and one minimum, then an adjacent pair of maxima and minima of $\tilde{h}_{\partial\hat{V}}$ may be cancelled. Here the maximum corresponds to some saddle σ_1 of $\tilde{h}_{\partial\hat{V}}$ and the minimum corresponds to some saddle σ_2 of $\tilde{h}_{\partial\hat{V}}$. We may assume that there are no maxima or minima of $\tilde{h}_{K_2}$ in the cylindrical portion P of $\hat{V}$ between B_{σ_1} and B_{σ_2} (as defined in the remark above), by isotoping any such maximum or minimum up or down as necessary. When the maximum and minimum of $\tilde{h}_{\partial\hat{V}}$ are cancelled, P is levelled. Any subarc of K_2 entering into B_{σ_1} from P must have a maximum before any minima and a subarc of K_2 entering into B_{σ_2} from P must have a minimum before any maximum. By choosing a fixed subarc of K_2 between a maximum in B_{σ_1} and a minimum in B_{σ_2}, this subarc may be levelled, to cancel a maximum and minimum when P is levelled. No other critical points of $\tilde{h}_{K_2}$ need be affected.

Each pair of maxima and minima of $\tilde{h}_{\partial\hat{V}}$ that is cancelled corresponds to a pair of maxima and minima of K_1. By induction, we see that $b(K_2) \leq b(K_1 \# K_2) - b(K_1) + 1$. This proves that $b(K_1 \# K_2) \geq b(K_1) + b(K_2) - 1$.

The other inequality follows by considering a copy of K_2 in bridge position realizing $b(K_2)$ lying below a copy of K_1 in bridge position realizing $b(K_1)$ and taking the connected sum. □

References

[1] *Über eine numerische Knoteninvariante*,
H. Schubert, Math. Z. 61 (1954) 245-288

figure=add8.ps

fig. 1

figure=add5.ps

fig. 2

figure=add9.ps

fig. 3

Department of Math and CS, Emory University, Atlanta, GA 30322
E-mail address: `jcs@mathcs.emory.edu`