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: $b(K) \geq k \cdot (b(J))$. 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 main result may be recovered by a much shorter proof. This shorter proof grew out of an endeavour to recast the problem within the framework of the thin position of a knot. This framework turns out to be far more refined than necessary. The proof here does not employ the notion of thin position. It does, however, rely 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. In this way it differs dramatically from Schubert's proof. It 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) and their induced foliations. This streamlines the terminology and the complexity of the argument. Schubert's proofs of the results reproven here involve 25 pages containing 15 lemmas which involve a consideration of up to three cases.

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 and for helpful conversations. 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* <u>*bridge position*</u> *with respect to h. The* <u>*bridge number of K*</u>*, $b(K)$, 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,*

1991 *Mathematics Subject Classification.* 57N10.
Research partially supported by NSF grant DMS-9803826.

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 $\hat{V}$ 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. (It is also called the wrapping number.)*

In the special case in which the index of the pattern is 1, this construction yields the connected sum of J and L, and V is called a swallow-follow torus.

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, D_1 cobounds a 3-ball B with a disk $\tilde{D}_1 \subset L - T$, 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 inessential saddle in $\mathcal{F}_T$ that is innermost in 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 $a_1, \ldots, a_n$ 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. Alter T as in Lemma 1 so that there is an inessential saddle σ 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 horizontally 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 to a saddle in this way, since $\chi(T) = 0$.

Definition 5. *Let σ, c_1, c_2, c_3 be as above. We may assume that c_1 and c_2 are in the same level surface L of h. Then since L is a sphere, c_1 and c_2 cobound an annulus in L. If a collar of $c_1 \cup c_2$ in this annulus is contained in V, then σ is a nested saddle.*

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. Suppose that there are nested saddles in $\mathcal{F}_T$.

Claim: There are also saddles in $\mathcal{F}_T$ that are not nested.

Let σ be the highest saddle in $\mathcal{F}_T$. For c_1, c_2, L as above, let $\hat{D}_1$, $\hat{D}_2$ be the (disjoint) disks bounded by c_1, c_2 in L. As σ is the highest saddle in $\mathcal{F}_T$, any curve in $T \cap interior(\hat{D}_i)$ bounds a disk lying above L. This implies that $\hat{D}_i$ is isotopic to a disk whose interior is disjoint from T, i.e., lies either entirely in V or entirely in $S^3 - V$. Since c_i is parallel to s_i, c_i is essential in T. Furthermore, since V is knotted, T is incompressible, whence c_i is essential in the closure of $S^3 - V$. This implies that $\hat{D}_i$ must be isotopic to a disk whose interior lies entirely in V (in particular, $\hat{D}_i$ is a meridian disk). Thus σ is not nested.

If there are both saddles that are nested and saddles that are not nested, then there must be an "adjacent" pair σ_1, σ_2 of essential saddles in T 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 level disk (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 T.

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 $\hat{B}_2$ together with D. By the proof of Lemma 1, we may assume that $\hat{B}_2$ contains ∞ and that $s_2^2 \subset \hat{B}_2$. We may thus shrink horizontally and lower $(K \cup T) \cap B$ as in the proof of Lemma 2. The difference is that here $K \cup T$ meets D along its horizontal portion. As $(K \cup T) \cap B$ is shrunk horizontally, the horizontal portion of D is lowered while remaining horizontal. The portion of B lying above $h^{-1}(h(\sigma_1))$ is shrunk horizontally as necessary. In the end, a product neighborhood below the original horizontal portion of D ends 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 2. *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. It follows that if $L = h^{-1}(r)$ for some regular value r, then $V \cap L$ consists of disks. More specifically, let σ_1, …, σ_n be the saddles in $\mathcal{F}_T$, and let $L_i = h^{-1}(h(\sigma_i))$. Recall that each saddle σ corresponds to a maximum or minimum m_σ of h_T. Between the level surfaces $h^{-1}(h(\sigma))$ and $h^{-1}(h(m_\sigma))$ lies a portion B_σ of V that is a 3-ball. Here $L_1 \cup \cdots \cup L_n$ cuts V into B_{σ_1}, …, B_{σ_n} and vertical cylinders.*

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 as described in Remark 2. 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 as in Remark 2. For distinct saddles σ_i and σ_j, B_{σ_i} and B_{σ_j} are 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 1). By renumbering, if necessary, we may assume that $b(K_1) \geq b(K_2)$. Then we still obtain a Morse function on (S^3, K_1) as above. Furthermore, if, for each maximum of T and σ, B_σ, $\tilde{D}_1$, $\tilde{D}_2$ as above, $|\tilde{D}_i \cap K| \geq 2$ for $i = 1, 2$, then B_σ contains at least two maxima of K. Hence $b(K) \geq 2 \cdot b(K_1) \geq b(K_1) + b(K_2) - 1$. Thus we may assume that there is a meridian disk contained in a level surface of h that intersects K once.

Recall from Remark 2 that V is comprised of $B_{\sigma_1}, \ldots, B_{\sigma_n}$ and vertical cylinders. In a vertical cylinder, a meridian disk contained in a level surface that intersects K once may be used to move all critical points upwards or downwards and out of the cylinder. Thus the intersection of K with the cylinder becomes a monotone arc and the number of critical points of h_K is unchanged.

In B_{σ_1}, assume that $|\tilde{D}_1 \cap K| = 1$. We may assume that σ_1 corresponds to a maximum of h_T. Let α be the subarc of $B_{\sigma_1} \cap K$ that connects $\tilde{D}_1 \cap K$ to the closest maximum of K in B_{σ_1}. We may assume that this maximum is the highest maximum of K in B_{σ_1}. Then consider a disk E in B_{σ_1} for which ∂E consists of four subarcs: α, a_1, a_2, a_3, where a_1 and a_3 are horizontal arcs connecting the endpoints of α to $\partial B_{\sigma_1} \subset T$, and a_2 is an arc in ∂B_{σ_1} connecting the other endpoints of a_1 and a_3, that runs over the maximum of ∂B_{σ_1}, and has no other critical points. We further require that $E \cap T = a_2$.

Claim: After an isotopy that does not change the number of critical points of h_K, $E \cap K = \alpha$.

Let $p_1, \ldots, p_k$ be the points in $E \cap K - \alpha$ with p_n the highest such point. A small monotone subarc β of K containing p_n may be replaced by a monotone arc β' that begins at one endpoint of $K - \beta$, travels parallel to E until it almost reaches ∂B_{σ_1}, then circles around to the other side of E along ∂B_{σ_1} and travels parallel to E on the other side of E until it meets the other endpoint of $K - \beta$. See fig. 4. The result is isotopic to K and has the same number of critical points as K, yet one fewer intersection with E. The Claim follows by induction.

fig. 4

Now $B_{\sigma_1} \cap K$ may be isotoped horizontally and downward, so that after the isotopy this intersection consists of one arc with exactly one critical point. In the vertical solid cylinder meeting B_{σ_1} at $\tilde{D}_2$, $|\tilde{D}_2 \cap K| = 1$ and $\tilde{D}_2$ may be used to isotope K so that all critical points are moved to B_{σ_2} (after relabelling) and K intersects the solid cylinder in a single monotone arc. Here σ_2 corresponds to a minimum, but an identical argument shows how proceed. After a finite number of iterations of this procedure, $B_{\sigma_i} \cap K$ consists of a single arc with one critical point for $i = 1, \ldots, n-1$ and K intersects all cylindrical portions of V in monotone arcs. Then, (since σ_n corresponds to a minimum) ∂B_{σ_2} cuts K into $K_1 - (\textit{subarc containing a minimum})$ and $K_2 - (\textit{subarc containing a maximum})$. 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

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

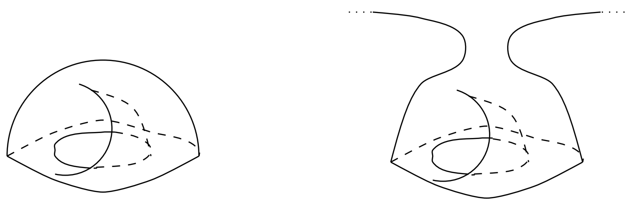

fig. 1

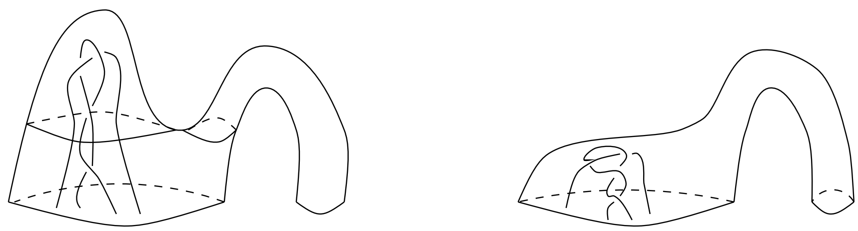

fig. 2

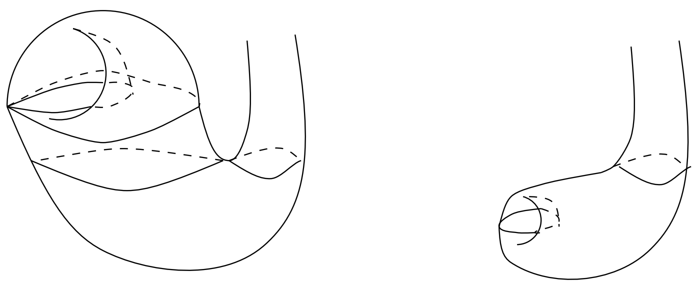

fig. 3

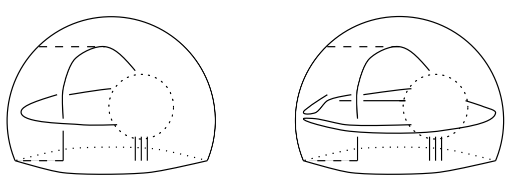

fig. 4