Optimal Pants Decompositions and Shortest Freely Homotopic Loops on an Orientable Surface

Éric Colin de Verdière*   Francis Lazarus†

1 Introduction

Let $\mathcal{M}$ be a compact orientable combinatorial surface of genus g with b boundaries. A *pants decomposition* of $\mathcal{M}$ is a maximal set of pairwise disjoint, non-isotopic, essential loops on $\mathcal{M}$; a loop being *essential* if it is simple and neither contractible nor homotopic to a boundary of $\mathcal{M}$. A pants decomposition is made of $3g-3+b$ loops and cuts $\mathcal{M}$ into *pairs of pants*, *i.e.*, spheres with three boundaries (see [4]).

We describe a conceptually simple, polynomial, iterative scheme which takes a given pants decomposition and outputs a shorter homotopic pants decomposition. We prove that, at the end of the process, each loop is a shortest loop in its homotopy class (in this paper, we consider homotopy of loops without basepoint, *i.e.*, free homotopy). In particular, the resulting decomposition is *optimal* in the sense that it is as short as possible among all homotopic decompositions.

Furthermore, given a simple, essential loop ℓ, it is not difficult to extend ℓ to a pants decomposition of $\mathcal{M}$ (see [3]). This decomposition, after optimization, contains a shortest loop homotopic to ℓ which is simple. Even the existence of such a simple loop is non-obvious.

The problem of optimizing a pants decomposition was raised in the conclusion of [3]; to our knowledge, we present the first algorithm which solves it. It also somehow extends [5] to more general surfaces. This is a natural extension of our former paper [1] where we treat the case of optimal simple loops in a given class of homotopy with fixed basepoint as opposed to free homotopy.

*Laboratoire d'informatique de l'École normale supérieure, UMR 8548 (CNRS), Paris, France. Eric.Colin.de.Verdiere@ens.fr

†Laboratoire IRCOM-SIC, UMR 6615 (CNRS), Poitiers, France. Francis.Lazarus@sic.univ-poitiers.fr

2 Framework and Result

The framework we use in this paper is very close to the one used in [1], see this paper for details. The surface $\mathcal{M}$ is assumed to be a polyhedral 2-manifold, whose edges have positive weights. Let G be the vertex-edge graph of $\mathcal{M}$, and G^* be its dual (embedded into $\mathcal{M}$). We consider sets of disjoint, simple, piecewise linear (PL) curves drawn on $\mathcal{M}$ that intersect G^* in an *admissible* way (*i.e.*, the intersections are generic). Throughout this paper, we always assume admissibility. If a curve crosses the edges $e_1^*, \dots, e_k^*$ of G^*, then its *length* is defined to be the sum of the weights of $e_1, \dots, e_k$. This notion of length coincides with the usual length if we retract all curves on G.

Let s^{init} be a pants decomposition of $\mathcal{M}$ to be optimized[1]. To simplify the computation and the proof of correctness, we first augment s^{init} to form a *doubled pants decomposition*, which we call s. Is is obtained by taking a copy of each loop in s^{init} and of each boundary[2] of $\mathcal{M}$, slightly translated, in the same homotopy class, such that s is still a set of pairwise disjoint simple loops. $s = (s_1, \dots, s_N)$ is thus composed of $N = 6g-6+3b$ loops. A loop of s or a boundary of $\mathcal{M}$ and its translated copy are called *twins*. For a loop s_j in s, the connected component of $\mathcal{M} \setminus \{s \setminus s_j\}$ that contains s_j is a pair of pants, and one of its three boundaries is the twin of s_j. We note $\mathcal{P}_j$ this pair of pants.

Definition 1 *An* Elementary Step $f_j(s)$ *consists in replacing the* j*th loop* s_j *by a shortest simple homotopic loop in* $\mathcal{P}_j$*. A* Main Step $f(s)$ *is the application of* $f = f_N \circ f_{N-1} \circ \dots \circ f_2 \circ f_1$ *to*

[1] In fact, we allow s^{init} to be a decomposition of $\mathcal{M}$ with pairs of pants and annuli, as it must be the case if $\mathcal{M}$ is a torus or a cylinder. This technicality does not change anything for the rest of the paper.

[2] This to allow shortening of loops that would be homotopic to a boundary.

s. These operations transform a doubled pants decomposition into another one, keeping the homotopy class of the decomposition.

Here is our main theorem:

Theorem 2 *Let s^0 be a doubled pants decomposition of $\mathcal{M}$, and let $s^{n+1} = f(s^n)$. For some $m \in \mathbb{N}$, s^m and s^{m+1} have the same length and, in this situation, s^m is a doubled pants decomposition homotopic to s^0 made of loops which are individually as short as possible among all loops in their (free) homotopy class. In particular, s^m is an optimal doubled pants decomposition of $\mathcal{M}$.*[3]

Since it is easy to extend a simple loop to a pants decomposition, and since a pants decomposition is made of simple loops, we have:

Corollary 3 *Let ℓ be a simple loop in $\mathcal{M}$. There exists a* simple *loop ℓ' homotopic to ℓ which is as short as possible among all loops homotopic to ℓ.*

The following section aims at proving Theorem 2. Note that Lemma 7 explains how to perform algorithmically the computations of f_i.

3 Proof of Theorem 2

Let π be the projection from the universal cover $\tilde{\mathcal{M}}$ of $\mathcal{M}$ onto $\mathcal{M}$. In this section, we fix $i \in [1, N]$; let s be a doubled pants decomposition, and let t_i be a loop which is homotopic to s_i and as short as possible among all loops homotopic to s_i. We may assume that no lift of t_i self-intersects in $\tilde{\mathcal{M}}$ (see below for the definition of a lift). This fact, which is not trivial, will be used in the proof of Proposition 9. Our goal is to prove that, after a finite number of steps, the ith loop of the doubled pants decomposition has the same length as t_i.

3.1 Lifts and translations in $\tilde{\mathcal{M}}$

Let ℓ be a loop on $\mathcal{M}$. We view ℓ as a 1-periodic mapping from $\mathbb{R}$ into $\mathcal{M}$. A *lift* of ℓ is a mapping $\tilde{\ell} : \mathbb{R} \to \tilde{\mathcal{M}}$ such that $\pi \circ \tilde{\ell} = \ell$. A *part* of a lift $\tilde{\ell}$ is the restriction of $\tilde{\ell}$ to an interval of the form $[a, a+1)$.

Let $\tilde{\ell}_1$ be a part of a lift $\tilde{\ell}$. Let v be a point in $\tilde{\mathcal{M}}$, and let β be a path from the source of $\tilde{\ell}_1$ to v. Consider the target v' of the lift of $\pi(\beta)$ starting at the target of $\tilde{\ell}_1$. It is readily seen that v' does not depend on the path β, nor on the part $\tilde{\ell}_1$ of $\tilde{\ell}$ chosen. We have $\pi(v) = \pi(v')$; intuitively, v' is the translated of v by $\tilde{\ell}$. We define $\tau_{\tilde{\ell}}(v)$ to be v'.

Let s_j and $s_{j'}$ be two twins of s. Let $(s_j^\alpha)^{\alpha \in \mathbb{N}}$ be an enumeration of the lifts of s_j in $\tilde{\mathcal{M}}$. By [2, Lemma 2.4], s_j and $s_{j'}$ bound a cylinder in $\mathcal{M}$. It follows that, for each $\alpha \in \mathbb{N}$, s_j^α is one boundary of an infinite strip which contains no lift of s in its interior, and is bounded from the other side by a lift of $s_{j'}$. We call $s_{j'}^\alpha$ this other boundary.

Let $j \in [1, N]$. For a lift $\tilde{t}_i$ of t_i, $\tau_{\tilde{t}_i}$ induces a permutation σ_j of $\mathbb{N}$ as follows: the image by $\tau_{\tilde{t}_i}$ of s_j^α is also a lift of s_j, which we call $s_j^{\sigma_j(\alpha)}$. The σ_j's, which depend on $\tilde{t}_i$ and on the enumeration of the lifts of s_j, will remain fixed in the rest of this paper.

3.2 Crossing words

Let A be the set of *symbols* of the form k^α or $\bar{k}^\alpha$, where $k \in [1, N]$ and $\alpha \in \mathbb{N}$. The set A^* of *words* on A is the set of finite sequences of elements in A. Let $\tilde{p}$ be a path in $\tilde{\mathcal{M}}$; $\tilde{p}$ crosses the lifts s_k^α (for $k \in [1, N]$ and $\alpha \in \mathbb{N}$) at a finite number of points. We walk along $\tilde{p}$ and, at each crossing encountered with a lift s_k^α of s, we write the symbol k^α or $\bar{k}^\alpha$, according to the orientation of the crossing (with respect to a fixed orientation of $\tilde{\mathcal{M}}$). The resulting element of A^* is called the *crossing word* of $\tilde{p}$ with s, and denoted by $s/\tilde{p}$.

Lemma 4 *Let $a^1 < a^2$ be two real numbers such that exactly one crossing occurs between all lifts of s and $\tilde{t}_i|_{[a^1,a^2)}$. For $k = 1, 2$, let $w^k = s/\tilde{t}_i|_{[a^k,a^k+1)}$.*

If $w^1 = j^\alpha.w$ (resp. $w_1 = \bar{\jmath}^\alpha.w$), then $w^2 = w.j^{\sigma_j(\alpha)}$ (resp. $w^2 = w.\bar{\jmath}^{\sigma_j(\alpha)}$).

Let $w \in A^*$. We define the relation $\sim$ to be the equivalence relation generated by $j^\alpha.w \sim w.j^{\sigma_j(\alpha)}$ and $\bar{\jmath}^\alpha.w \sim w.\bar{\jmath}^{\sigma_j(\alpha)}$ (for any j and α). Let $[A^*]$ be A^* quotiented by the relation $\sim$. If $w \in A^*$, we denote by $[w]$ its equivalence class in $[A^*]$.

[3]**Remark.** The proof of Theorem 2 extends to the case where we consider the real *length* of PL loops drawn on $\mathcal{M}$ (and not on its vertex-edge graph), provided that the suitable definition of a crossing is used: we have to take into account that two loops can partly overlap without crossing.

Let $\tilde{t}_i^1$ be a part of $\tilde{t}_i$. It follows from the previous lemma that $[s/\tilde{t}_i^1]$ does not depend on $\tilde{t}_i^1$; hence we define $[s/\tilde{t}_i]$ to be their common equivalence class in $[A^*]$.

Let $j \in [1, N]$, and let $[w] \in [A^*]$. The *j-reductions* of $[w]$ are defined as follows. If w has the form $w_1 j^\alpha \bar{\jmath}^\alpha w_2$ or $w_1 \bar{\jmath}^\alpha j^\alpha w_2$, then we say that $[w]$ *j-reduces* to $[w_1 w_2]$; if w has the form $j^\alpha w_1 \bar{\jmath}^{\sigma_j(\alpha)}$ or $\bar{\jmath}^\alpha w_1 j^{\sigma_j(\alpha)}$, then we also say that $[w]$ *j-reduces* to $[w_1]$. Obviously, this definition does not depend on the particular choice of the word w in $[w]$.

A *reduction* is a j-reduction for some j. $[w]$ is *j-irreducible* (resp. *irreducible*) if it can be applied no j-reduction (resp. reduction).

Lemma and Definition 5 *Let $[w] \in [A^*]$. There is only one j-irreducible (resp. irreducible) element of $[A^*]$ which can be obtained from $[w]$ by successive j-reductions (resp. reductions). We define $g_j([w])$ (resp. $g([w])$ to be this word.*

3.3 Reducibility of $[s/\tilde{t}_i]$

Proposition 6 *$g([s/\tilde{t}_i]) = \varepsilon$, where ε is the class of the empty word in $[A^*]$.*

Proof. Let s_i' be a loop homotopic to s_i and slightly translated such that it does not cross any loop s_k. Let $\dot{s}_i$ and $\dot{t}_i$ be the restrictions of s_i' and t_i to $[0,1]$. There exists a path β joining $s_i'(0)$ to $t_i(0)$ such that the path $p := \beta.\dot{t}_i.\beta^{-1}.\dot{s}_i^{-1}$ is a null-homotopic loop in $\mathcal{M}$. We subdivide p into four paths $p_1 = \beta$, $p_2 = \dot{t}_i$, $p_3 = \beta^{-1}$, and $p_4 = \dot{s}_i^{-1}$. Let $\tilde{p} = \tilde{p}_1.\tilde{p}_2.\tilde{p}_3.\tilde{p}_4$ be a lift of p such that $\tilde{p}_2$ is on $\tilde{t}_i$.

It can be proved that $s/\tilde{p}$ is a parenthesized expression (it reduces to the empty word by successive removals of subwords of the form $j^\alpha \bar{\jmath}^\alpha$ and $\bar{\jmath}^\alpha j^\alpha$). Hence $g([s/\tilde{p}]) = \varepsilon$.

$\tilde{p}_1$ and $\tilde{p}_3^{-1}$ are parts of lifts of β, and $\tau_{\tilde{t}_i}(\tilde{p}_1)$ is equal to $\tilde{p}_3^{-1}$. Hence, if the kth symbol of $s/\tilde{p}_1$ is equal to j^α (resp. $\bar{\jmath}^\alpha$), then the kth symbol of $s/\tilde{p}_3^{-1}$ (which equals $s/\tilde{p}_3$ in reverse order) is $j^{\sigma_j(\alpha)}$ (resp. $\bar{\jmath}^{\sigma_j(\alpha)}$). Since $s/\tilde{p}_4$ is empty, it follows that $g([s/\tilde{p}_2]) = g([s/\tilde{p}])$. The left handside equals $g([s/\tilde{t}_i])$ and the right handside equals ε. □

3.4 Uncrossing the loops

Lemma 7 *Let $r = f_j(s)$. r_j is, in $\mathcal{P}_j$, a shortest loop homotopic to s_j.*

Proof (sketch). Let b_k, $k = 1, 2, 3$, be the boundaries of $\mathcal{P}_j$, such that b_1 is homotopic to s_j. Let p_1 (resp. p_2) be a shortest path between b_2 and b_3 (resp. b_1 and b_3); we can make these paths simple and disjoint. Let ℓ be a shortest loop homotopic to s_j in $\mathcal{P}_j$, and $\tilde{\ell}$ be a lift of ℓ in the universal cover of $\mathcal{P}_j$. By analyzing the way $\tilde{\ell}$ crosses the lifts of p_1 and p_2, we can prove that we can change ℓ to a loop ℓ', which is also a shortest loop homotopic to s_j in $\mathcal{P}_j$, but does not cross p_1 and crosses p_2 once.

Cut $\mathcal{P}_j$ along p_1 and p_2; for each pair of vertices corresponding to a single vertex of p_2 before cutting, compute a shortest path whose endpoints are this pair of vertices; take the shortest of these shortest paths. By the preceding paragraph, this path yields a shortest loop homotopic to s_j in $\mathcal{P}_j$, and it is simple. As a byproduct, this describes a way to compute $f_j(s)$. □

Let $r = f_j(s)$. If $k \neq j$, let r_k^α be equal to s_k^α. To get an enumeration of the lifts of r_j, we proceed as follows. Let $s_{j'}$ be the twin of s_j. Note that r_j and $s_{j'}$ bound a cylinder by [2, Lemma 2.4]. We let r_j^α to be the lift of r_j which bounds the lift of this cylinder whose other boundary is $s_{j'}^\alpha$. It follows that $\tau_{\tilde{t}_i}(r_j^\alpha)$ is equal to $r_j^{\sigma_j(\alpha)}$ (in other words, the permutation σ_j remains unchanged).

Lemma 8 $g_j([r/\tilde{t}_i]) = g_j([s/\tilde{t}_i])$.

Proof (sketch). Let $[r/\tilde{t}_i]_j$ and $[s/\tilde{t}_i]_j$ be obtained by deleting j-symbols from $[r/\tilde{t}_i]$ and $[s/\tilde{t}_i]$. Since r and s only differ in their jth loop, these two words are identical. Consider two consecutive symbols σ_1 and σ_2 in $[u/\tilde{t}_i]_j$, where u stands for either r or s. These two symbols are replaced in $[u/\tilde{t}_i]$ by an expression $\sigma_1 w_j \sigma_2$, where w_j is a word on j-symbols. We only need to show that w_j reduces (with parenthesized reductions) to a same expression for $u = r$ and $u = s$. This obviously implies the lemma. The proof uses the fact that $u_{j'}$ $(= s_{j'})$ and u_j bound a cylinder in $\mathcal{P}_j$, and this cylinder is crossed by no other loops of u. □

Proposition 9 *We can replace t_i by a loop t_i' (homotopic to t_i, no longer than t_i, and such that its lifts are simple) so that $[r/\tilde{t}_i'] = g_j([s/\tilde{t}_i])$ for some lift $\tilde{t}_i'$ of t_i'.*

Proof. By Lemma 8, we may only consider the case where $[r/\tilde{t}_i']$ is j-reducible; this implies that there is a disk D in $\tilde{\mathcal{M}}$ bounded by an arc $\tilde{r}_j^{ab}$ of a lift $\tilde{r}_j$ of r_j, and an arc $\tilde{t}_i^{ab}$ of $\tilde{t}_i$ with the same endpoints a and b.

D intersects $\tilde{t}_i$ in a set of pairwise disjoint arcs with endpoints on $\tilde{r}_j^{ab}$ (recall $\tilde{t}_i$ is simple). Consider an innermost such arc $\tilde{t}_i^{cd}$, *i.e.*, such that it sustains a subarc $\tilde{r}_j^{cd}$ of $\tilde{r}_j^{ab}$ that does not intersect $\tilde{t}_i$.

If $\tilde{t}_i^{cd}$ were shorter than $\tilde{r}_j^{cd}$, we could shorten r_j as follows: in $\tilde{\mathcal{M}}$, replace the part $\tilde{r}_j^{cd}$ of $\tilde{r}_j$ by a path with the same endpoints going along $\tilde{t}_i^{cd}$, and project it onto $\mathcal{M}$. The resulting loop, r'_j, is shorter than r_j; moreover, no lift of any loop other than t_i can cross D, so the projection $\pi(D)$ lies entirely in $\mathcal{P}_j$. It follows that r'_j is homotopic *in* $\mathcal{P}_j$ to r_j, while being shorter; this contradicts Lemma 7.

We modify $\tilde{t}_i$ as follows: replace the part $\tilde{t}_i^{cd}$ of $\tilde{t}_i$ by a path with the same endpoints going along $\tilde{r}_j^{cd}$, on the other side of $\tilde{r}_j^{cd}$ (to remove the two crossings). The projection t'_i of the resulting path is a loop homotopic to t_i, and no lift of this new loop self-intersects in $\tilde{\mathcal{M}}$. It cannot be longer than t_i by the preceding paragraph, hence t'_i is a shortest loop homotopic to s_i whose lifts are simple. Moreover, $[r/\tilde{t}'_i]$ is deduced from $[r/\tilde{t}_i]$ by a j-reduction. We finish the proof by induction. □

3.5 Conclusion of the proof

Lemma 10 *Assume t_i does not cross any loop of s; let $\mathcal{P}$ be the pair of pants delimited by s in which t_i is. Then one of the boundaries of $\mathcal{P}$ is homotopic,* in $\mathcal{P}$, *to t_i.*

PROOF. Omitted in this abstract. □

Lemma 11 *Assume that $[s/\tilde{t}_i] = \varepsilon$. Let $r = f^2(s)$. Then r_i and its twin have the same length as t_i.*

PROOF. By Lemma 10, t_i is inside a pair of pants bounded by some s_k which is either s_i or its twin. By Proposition 9, we may replace t_i by t'_i such that $s' := f_{k-1} \circ \ldots \circ f_1(s)$ does not cross t'_i, and (in fact) that t'_i is in a pair of pants bounded by s'_k. Hence, by Lemma 7, the kth loop of $f_k(s')$ has the same length as t_i. After one more iteration of f the same is true for the twin of the kth loop. □

PROOF OF THEOREM 2. Fix i; let t_i^0 be a shortest loop homotopic to s_i^0. By Propositions 9 and 6, one can construct a sequence $(t_i^n)_{n\in\mathbb{N}}$ of shortest homotopic loops such that the length of $[s^n/t_i^n]$ strictly decreases. Then for some n, $[s^n/t_i^n] = \varepsilon$. By Lemma 11, s_i^{n+2} has the same length as t_i^0. Hence the length of s^n becomes stationary. It remains to prove that the lengths remain unchanged once s^n and s^{n+1} have the same lengths. □

4 Complexity

We present a sketch of the complexity analysis (which is similar to and simpler than the one in [1]). Let n be the number of edges of $\mathcal{M}$, g its genus and b its number of boundaries. Let α be the longest-to-shortest edge ratio of $\mathcal{M}$. Let S be a combinatorial doubled pants decomposition of $\mathcal{M}$ composed of $N = O(g+b)$ loops, and μ be the maximal multiplicity of any vertex of $\mathcal{M}$ in a loop of S. Hence the number of edges of a loop at the beginning of the algorithm is $O(\mu n)$, and, since loops can only get shorter in length, their maximal number of edges is $O(\alpha\mu n)$. We can prove that the lengths of the crossing words is $O((g+b)\alpha\mu^2 n)$, and compute the time spent by an Elementary Step, using Dijkstra's algorithm and the proof of Lemma 7. Finally:

Theorem 12 *This algorithm computes an optimal pants decomposition homotopic to S in $O(\mu^4\alpha^3(g+b)^2 n^3 \log \mu\alpha n)$ time.*

References

[1] É. Colin de Verdière and F. Lazarus. Optimal system of loops on an orientable surface. In *IEEE Symp. Found. Comput. Sci.*, pages 627–636, 2002.

[2] D. Epstein. Curves on 2-manifolds and isotopies. *Acta Mathematica*, 115:83–107, 1966.

[3] J. Erickson and S. Har-Peled. Optimally cutting a surface into a disk. In *Proc. 18th Annu. ACM Symp. Comput. Geom.*, pages 244–253, 2002.

[4] A. Hatcher. Pants decompositions of surfaces. http://www.math.cornell.edu/˜hatcher/Papers/pantsdecomp.pdf, 2000.

[5] J. Hershberger and J. Snoeyink. Computing minimum length paths of a given homotopy class. *Comput. Geom. Theory Appl.*, 4:63–98, 1994.