STABLE COMMUTATOR LENGTH

DANNY CALEGARI

ABSTRACT. This is an introduction to the theory of stable commutator length and quasimorphisms. It is not meant to be a comprehensive survey or bibliography, nor are results presented in the greatest possible generality or with complete proofs. Rather the emphasis is on examples and clarity of exposition, and on the underlying fundamental principles.

CONTENTS

1. SURFACES

1.1. **Motivation.** Let X be a space. Classically, X is studied by looking at maps from 1-manifolds to X. See Figure 1.

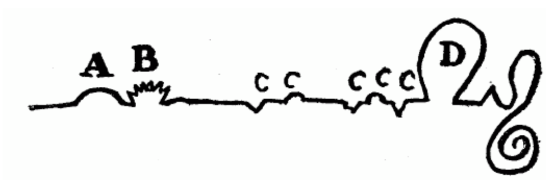

FIGURE 1. (from *Tristram Shandy* by Lawrence Sterne)

In the Riemannian world, we get geodesics, Jacobi fields, Morse theory, etc. In the topological world, we get the fundamental group(oid), covering spaces, etc.

It is a relatively recent idea to study X by looking at maps from surfaces to X (this is morally a kind of *complexification* of geometry). See Figure 2.

In the Riemannian world, we get minimal surfaces, (pseudo-) holomorphic curves, string theory, etc. We are concerned with the topological side of this story.

Date: April 30, 2015.

FIGURE 2. (after Frank Morgan and Sir John Tenniel)

The theory of stable commutator length and stable genus, and the dual theory of quasimorphisms and bounded 2-dimensional cohomology, is concerned with *quantitative* topological questions about maps from surfaces to spaces; questions like:

(1) What is the simplest spanning surface of a knot?
(2) Which Seifert fibered spaces admit taut foliations?
(3) How small can b^2/σ be for a smooth, simply connected 4-manifold?

1.2. **The classification of surfaces.** We are interested here and in the sequel in compact, oriented surfaces, possibly disconnected, possibly with boundary. The following operations on surfaces are fundamental:

(1) passing to finite covers;
(2) disjoint union; and
(3) cut-and-paste

The first two need no explanation, except to emphasize that the covers in question may be disconnected, even if the base surface is connected. By *cut-and-paste* we mean repeated applications of the following operation ("paste") and/or its inverse ("cut"): starting with a compact oriented surface S and two disjoint loops ∂^+, ∂^- in the boundary, we form a compact oriented surface S' by gluing ∂^+ to ∂^- by an orientation-reversing homeomorphism. A *compression* is performed on a surface S first by cutting along an embedded loop γ, then by pasting in two disjoint disks along the resulting boundary components. *Essential cut-and-paste* is that in which neither boundary component $\partial^\pm$ bounds a disk component of S.

Connected, compact, oriented surfaces are classified by genus g and number of boundary components b. These can be arbitrary non-negative integers. Surfaces can also be filtered by *Euler characteristic* χ; this is a single integer taking values in $(-\infty, 2]$, and there are only finitely many connected, compact oriented surfaces whose Euler characteristic is any given value. The relationship between these invariants is

$$\chi = 2 - 2g - b$$

Only the sphere and the disk have positive Euler characteristic (2 and 1 respectively); only the annulus and torus have zero Euler characteristic. All other compact, connected, oriented surfaces have strictly negative Euler characteristic.

Euler characteristic is well-defined for compact surfaces which are not necessarily connected, and is additive under disjoint union and multiplicative under finite covers (also not assumed to be connected). It is also additive under cut-and-paste. For S a connected surface, define $\chi^-(S) := \min(\chi(S), 0)$ and for S compact but not necessarily connected, define $\chi^-(S) = \sum_i \chi^-(S_i)$ where the sum is taken over the components S_i of S. The function χ^- is still multiplicative under covers, but is no longer additive under cut-and-paste unless one excludes sphere and disk components. Genus can be extended to disconnected surfaces, by defining $\text{genus}(S) := \sum_i \text{genus}(S_i)$ where S_i are the components of S.

A compact, oriented surface without spheres, disks, annuli or tori (equivalently, for which every component has negative Euler characteristic) is said to be *hyperbolike*. The quantity $-\chi^-$ is a good measure of complexity for hyperbolike surfaces, and there are only finitely many hyperbolike surfaces for which $-\chi^-$ is any given value. For S a compact surface possibly with boundary, a *hyperbolic structure* on S is a Riemannian metric of constant curvature -1 and geodesic boundary. A hyperbolike surface admits a unique hyperbolic structure in every conformal class of metric. Furthermore, a cut-and-paste decomposition of a hyperbolike surface into hyperbolike pieces can be realized up to isotopy by a decomposition of hyperbolic surfaces into hyperbolic pieces.

Every hyperbolike surface S can be decomposed into $-\chi^-(S)$ pairs of pants, usually in many ways.

1.3. **Finite covers.** The theory of finite covers of surfaces is surprisingly rich, and much is still not known. For the moment we discuss two interactions of cut-and-paste with finite covers.

1.3.1. *Virtual compression.* Let S be a compact, oriented surface mapping to a space X. Suppose the map is not homotopically essential, so that there is some homotopy class of essential loop $\gamma : S^1 \to S$ whose image in X is homotopically trivial.

Scott [9] proved the following theorem:

Theorem 1.1 (Scott). *If $\gamma : S^1 \to S$ is essential then there is a finite cover $\hat{S}$ of S and a lift $\hat{\gamma} : S^1 \to \hat{S}$ which is homotopic to an embedding.*

Since $\hat{\gamma} : S^1 \to \hat{S}$ is homotopic to an embedding, and the image of $\hat{\gamma}$ in X under $\hat{S} \to X$ is still homotopically trivial, the surface $\hat{S}$ can be compressed along $\hat{\gamma}$, thereby reducing its complexity.

1.3.2. *Virtual cut-and-paste.* A degree n oriented cover of S^1 is classified by a conjugacy class in the symmetric group S_n corresponding to the monodromy of the cover; equivalently, it is given by a partition of n into positive integers, which are the degrees of the covering map on the components.

Let S be a compact, oriented surface and suppose there are disjoint oriented 1-manifolds ∂^-, ∂^+ in ∂S mapping to S^1 under covering maps of the same degree which are respectively orientation-reversing and -preserving. We would like to find a finite cover $\hat{S} \to S$ for which the preimages $\hat{\partial}^-$ and $\hat{\partial}^+$ are homeomorphic by an orientation-reversing homeomorphism compatible with the covering maps to γ. This can be achieved if S is hyperbolike:

Lemma 1.2. *Let S be hyperbolike, and let $\Gamma \subset \partial S$ be a submanifold. Let $\Gamma \to S^1$ be a covering map (not necessarily oriented) in which some components might map with positive degree and others with negative degree. Then there is an integer N and a finite cover $\hat{S} \to S$ so that in the induced covering $\hat{\Gamma} \to S^1$ every component maps with (unoriented) degree N.*

Proof. Since S is hyperbolike, we can first pass to a cover if necessary in which every component of S has strictly positive genus. If S is connected and hyperbolike with positive genus, it has a connected two-fold cover in which each boundary component has two connected preimages, each mapping to it with degree one. Thus by replacing S by a finite cover if necessary, we can assume that for each component S_i of S the boundary components of ∂S_i in Γ come in pairs $\gamma_{i,j}$, $\gamma'_{i,j}$ which each map to S^1 with the same (unoriented) degree $n_{i,j}$. Now, let N be the lowest common multiple of the $n_{i,j}$. We define a homomorphism $\phi : H_1(S) \to \mathbb{Z}/N\mathbb{Z}$ by setting it equal to $n_{i,j}$ on $\gamma_{i,j}$ and $-n_{i,j}$ on $\gamma'_{i,j}$, and 0 on components of $\partial S - \Gamma$. This homomorphism is zero on each component of ∂S, and therefore extends to $H_1(S)$, by Mayer-Vietoris. Let $\hat{S}$ be the cover of S associated to ϕ. Then $\hat{S}$ has the desired properties. □

Setting $\Gamma = \partial^+ \cup \partial^-$ as above, we see that we can pass to a cover in which $\hat{\partial}^+$ and $\hat{\partial}^-$ can be glued up.

1.4. **Stable genus and stable commutator length.** Given a class of maps from surfaces to some fixed target space, it is natural to try to organize the maps by some measure of complexity, and to look for a representative minimizing the complexity. The question of which map has the least complexity has an *unstable* version and a *stable* version, where the stable version of the question considers a notion of complexity which behaves well under taking finite covers. One can also consider both absolute and relative versions of this question, depending on whether the domain surfaces have boundary or not.

1.4.1. *The absolute case.* Let X be a space, and let $\alpha \in H_2(X;\mathbb{Z})$ be given. A map $f : S \to X$ from a closed, oriented surface S to X *represents* α if the image $f_*[S]$ of the fundamental class of S in $H_2(S;\mathbb{Z})$ is equal to α.

Definition 1.3 (Genus)**.** The *genus* of α, denoted genus(α), is the least genus of a closed, oriented surface $S \to X$ representing α.

A map $f : S \to X$ from a closed, oriented surface S to X is *admissible* for α if $f_*[S] = n\alpha$ for some positive n. By abuse of notation, we write $n(S)$ for n, to emphasize how it depends on S.

Definition 1.4 (Stable genus)**.** The *stable genus* of α, denoted $\|\alpha\|$, is the infimum of $-\chi^-(S)/2n(S)$ over all surfaces $f : S \to X$ admissible for α.

The factor of 2 reflects the fact that $-\chi(S)$ is "almost" equal to $2 \cdot \text{genus}(S)$, at least for surfaces with fixed numbers of boundary components.

1.4.2. *The relative case.* Let X be a space, and $Y \subset X$ a subspace (note that any map $Y \to X$ can be replaced with a subspace up to homotopy by the mapping cylinder construction). Let $\alpha \in H_2(X, Y; \mathbb{Z})$ be a relative homology class.

A map of pairs $f : (S, \partial S) \to (X, Y)$ where S is a compact, oriented surface *represents* α if $f_*[S] = \alpha$ where $[S]$ is the fundamental class of S in $H_2(S, \partial S; \mathbb{Z})$.

Definition 1.5 (Relative genus)**.** The *genus* of α, denoted genus(α), is the least genus of a closed, oriented surface $(S, \partial S) \to (X, Y)$ representing α.

A map of pairs $f : (S, \partial S) \to (X, Y)$ is *admissible* for α if $f_*[S] = n\alpha$ for some positive n. By abuse of notation, we write $n(S)$ for n, to emphasize how it depends on S.

Definition 1.6 (Relative stable genus)**.** The *relative stable genus* of the pair α, denoted $\|\alpha\|$, is the infimum of $-\chi^-(S)/2n(S)$ over all surfaces $f : (S, \partial S) \to (X, Y)$ admissible for α.

Definition 1.7 (stable commutator length)**.** Let Y be a subspace of X, and let $\gamma \in H_1(Y)$ in the kernel of $H_1(Y) \to H_1(X)$ be given. The *stable commutator length* of γ, denoted $\text{scl}(\gamma)$, is the infimum of $\|\alpha\|$ over all $\alpha \in H_2(X, Y)$ with $\partial\alpha = \gamma$. Equivalently, it is the infimum of $-\chi^-(S)/2n(S)$ over all positive $n(S)$ and all surfaces $f : (S, \partial S) \to (X, Y)$ for which $f_*[\partial S] = n(S)\gamma$.

Example 1.8. An important special case is when L is an oriented 1-manifold, and $\Gamma : L \to X$ is some given homotopy class of map for which $\Gamma_*[L] = 0$ in $H_1(X)$. By replacing X with the mapping cylinder of Γ, we can convert Γ into an inclusion. Then $\text{scl}([L])$ depends on Γ, and by abuse of notation we write it as $\text{scl}(\Gamma)$.

The fact that the norms are defined in each case as the infimum of $-\chi^-(S)/2n(S)$ over infinitely many admissible surfaces S leaves open the possibility that this infimum might be *realized*. Surfaces realizing the infimum are important, and we call them *extremal*:

Definition 1.9. A map $f : S \to X$ is *extremal* for $\alpha \in H_2(X)$ if $\|\alpha\| = -\chi^-(S)/2n(S)$. A map $f : (S, \partial) \to (X, Y)$ is *extremal* for $\alpha \in H_2(X, Y)$ (resp. $\gamma \in H_1(Y)$) if $\|\alpha\| = -\chi^-(S)/2n(S)$ (resp. $\text{scl}(\gamma) = -\chi^-(S)/2n(S)$).

The property of being extremal for a specific class is preserved under taking finite covers and disjoint unions:

Lemma 1.10. *Suppose $f : S \to X$ or $f : (S, \partial S) \to (X, Y)$ is extremal for some class. Then for any finite cover $\pi : \hat{S} \to S$ the map $f \circ \pi : \hat{S} \to X$ or $f \circ \pi : (\hat{S}, \partial\hat{S}) \to (X, Y)$ is also extremal for the same class.*

Moreover, if $f_i : S_i \to X$ or $f_i : (S_i, \partial S_i) \to (X, Y)$ are all extremal for the same class, the map of disjoint unions $\coprod f_i : \coprod S_i \to X$ or $\coprod f_i : (\coprod S_i, \partial \coprod S_i) \to (X, Y)$ is extremal for this class.

Proof. This follows from the definition, since both $-\chi^-(S)$ and $n(S)$ are multiplicative under finite covers and additive under disjoint union. □

The property of being extremal passes to essential subsurfaces:

Lemma 1.11. *Let $f : (S, \partial S) \to (X, Y)$ be extremal for $f_*([S])$. Let $S' \subset S$ be an essential subsurface. Then the restriction $f|_{S'} : (S', \partial S') \to (X, f(\partial S'))$ is extremal for $f_*([S'])$.*

Proof. Suppose not, and let $f'' : (S'', \partial S'') \to (X, f(\partial S'))$ be more efficient than S'. Since S' is essential, after homotoping S' in S we can assume $S - S'$ is hyperbolike. By Lemma 1.2 we can find finite covers of S'' and of $(S - S')$ which glue together along a cover of $\partial S'$ to build a new surface which represents a multiple of $f_*([S])$ more efficiently than S does. □

Extremal surfaces can also be assembled from extremal pieces under amalgamation:

Lemma 1.12. *Let X, Z be subspaces of $X \cup Z$ with $X \cap Z = L$ where L is a closed 1-manifold. Let $f : (S, \partial S) \to (X, L)$ and $g : (S', \partial S') \to (Z, L)$ be extremal for classes α, α' in $H_2(X, L)$ and $H_2(Z, L)$ respectively so that $\gamma := f_*([\partial S]) = -g_*([\partial S'])$ in $H_1(L)$. Let α be the class in $H_2(X \cup Z)$ mapping to γ under the boundary map. Then there are finite covers $\hat{S}$ and $\hat{S}'$ of S and S' respectively which can be glued along their image in L in such a way as to build a new surface $\hat{S}''$ which is extremal for γ.*

Proof. Any surface representing a multiple of the class α can be cut along the preimage of L to produce surfaces mapping to X and Z. After a homotopy, these pieces can be assumed to be essential. It follows that all such surfaces are obtained by gluing pieces in this way, and the surface obtained by (virtually) gluing two extremal pieces will be extremal. □

An important property of extremal surfaces is that they are homotopically essential:

Proposition 1.13 (Extremal is injective)**.** *Let $f : S \to X$ or $f : (S, \partial S) \to (X, Y)$ be extremal. Then f is π_1-injective.*

Proof. Suppose that there is some essential $\gamma : S^1 \to S$ for which $f \circ \gamma : S^1 \to X$ is null-homotopic. By Scott's Theorem 1.1 we can find a finite cover $\hat{S}$ in which some lift $\hat{\gamma}$ of γ embeds. Then we can compress $\hat{S}$ along $\hat{\gamma}$, keeping it in the same (relative) homology class, thereby certifying that $\hat{S}$ is not extremal. By Lemma 1.10 the surface S is not extremal either. □

Stable genus and relative stable genus extend by linearity and continuity to pseudo-norms on $H_2(X)$ and $H_2(X, Y)$ respectively. Stable commutator length extends by linearity and continuity to a pseudo-norm on the kernel of $H_1(Y) \to H_1(X)$. Genus is *not* linear in general.

There is a very close relationship between stable genus and the 2-dimensional *Gromov norm*, as introduced by Gromov in his seminal paper [7]. If X is a space, there is an (L^1) norm on the real singular chain groups $C_i(X; \mathbb{R})$ of X, defined by

$$\left\| \sum t_i \sigma_i \right\|_1 := \sum |t_i|$$

If $\alpha \in H_i(X; \mathbb{R})$ is a homology class, define the *Gromov norm* $\|\alpha\|_1$ to be the infimum $\inf_C \|C\|_1$ over all i-cycles C representing the class α. It is a fact that for any space X

the Gromov norm is equal to 4 times the stable genus. This reflects the fact that it takes 4 triangles to build a "handle". This is proved by Thurston's method of *straightening simplices*.

1.5. **Circle bundles.** Different flavors of stable genus are related to each other functorially. Let X be a connected space, let $\alpha \in H_2(X;\mathbb{Z})$, and let $\phi \in H^2(X;\mathbb{Z})$ pair nontrivially with α. There is an oriented circle bundle E over X associated to ϕ (i.e. with Euler class equal to ϕ). Pick a basepoint $x \in X$ and let γ be the fiber of E over x. Since E is an oriented circle bundle, the monodromy of $\pi_1(X)$ on the homology of the fibers is trivial, and there is a spectral sequence for the homology of E whose E_2 term is $H_*(X;H_*(S^1))$. Then

$$0 \to H_2(X) \to H_1(\gamma) \to H_1(E) \to H_1(X) \to 0$$

is exact, and the map from $H_2(X) \to H_1(\gamma) = \mathbb{Z}$ takes a class to its pairing with ϕ. In particular, $H_1(\gamma) \to H_1(E)$ is trivial over $\mathbb{Q}$, and there is a (rational) class $\hat{\alpha}$ in $H_2(E,\gamma)$ mapping to α. The class $\hat{\alpha}$ is unique up to the subspace spanned by tori which are circle bundles over loops in X. These tori all have (stable) genus zero, and therefore the relative stable genus $\|\hat{\alpha}\|$ is well-defined, independent of the choice of representative.

Lemma 1.14 (Equal norms)**.** *With notation as above, there is equality* $\|\hat{\alpha}\| = \|\alpha\|$.

Proof. An inequality in one direction can be seen by composing a map $S \to E$ representing a multiple of $\hat{\alpha}$ with projection $E \to X$, and coning off the boundary components. In the other direction, if $S \to X$ (connected) represents a multiple $n \cdot \alpha$, there is a lift of $S - \text{point}$ to E whose (connected) boundary wraps n times around γ. Since $\chi(S - \text{point}) = \chi(S) - 1$, taking $n \to \infty$ proves the lemma. □

Remark 1.15. One subtlety of this correspondence is that α might be represented by an extremal surface, whereas $\hat{\alpha}$ will *never* be. For, if $S \to E$ represents $n\hat{\alpha}$, the projection $S \to X$ can be coned off along boundary components to $S' \to X$ and then there is a *strict* inequality

$$\|\hat{\alpha}\| = \|\alpha\| \le -\chi(S')/2n < -\chi(S)/2n$$

1.6. **Groups and commutators.** If G is a group, we can build a space X with $\pi_1(X) = G$. Conjugacy classes g in G correspond to homotopy classes of maps $\gamma : S^1 \to X$. A conjugacy class g can be written as a product of n commutators in G if and only if the corresponding loop $\gamma : S^1 \to X$ bounds a genus n surface; i.e. if and only if there is a compact, connected, oriented surface S with genus n and 1 boundary component, and a map $f : S \to X$ for which $f|_{\partial S}$ factors as $\gamma \circ h$ where $h : \partial S \to S^1$ is an orientation-preserving homeomorphism. The reason for this comes from the "standard" presentation for the fundamental group of S as a free group on $2n$ generators $a_1, b_1, \cdots, a_n, b_n$ in such a way that the boundary loop is conjugate to the product $\prod_i [a_i, b_i]$. Since S retracts to its 1-skeleton, any homomorphism between groups $\pi_1(S) \to G$ is induced by some map of spaces $S \to X$. Exhibiting g as a product $g = \prod_i [\alpha_i, \beta_i]$ determines a homomorphism, where $a_i \to \alpha_i$ and $b_i \to \beta_i$, and thereby a map $S \to X$ for which $\partial S \to X$ represents the homotopy class of γ.

The following definition explains the origin of the term *stable commutator length*:

Definition 1.16 (commutator length in groups)**.** Let G be a group, and let $[G,G]$ denote its commutator subgroup. The *commutator length* of $g \in [G,G]$, denoted $\text{cl}(g)$ is the smallest number of commutators whose product is g. The *stable commutator length* is the limit $\text{scl}(g) := \lim_{n\to\infty} \text{cl}(g^n)/n$.

Similarly, for $g_1, g_2, \cdots, g_k$ a finite collection of elements of G whose product is in $[G,G]$, define

$$\text{scl}\,(g_1 + g_2 + \cdots + g_k) = \lim_{n\to\infty} \text{cl}(g_1^n g_2^n \cdots g_k^n)/n$$

Proposition 1.17. *Let g_i be a finite collection of conjugacy classes in G, let $\gamma_i : S^1 \to X$ be the corresponding free homotopy classes of loops, and let $\Gamma : \coprod_i S^1 \to X$ denote the map of the union of circles, one for each index i. The product of the g_i is in $[G,G]$ if and only if $\Gamma_*[\coprod_i S^1] = 0 \in H_1(X)$ in which case there is an equality*

$$\text{scl}\,(g_1 + g_2 + \cdots + g_k) = \text{scl}(\Gamma)$$

For a proof, see [3], § 2.6. Note that the left hand side does not depend on the choice of space X with $\pi_1(X) = G$.

1.7. **Examples and basic properties.** In this subsection we briefly review some basic properties of stable genus, relative stable genus and stable commutator length, and give a few key examples.

1.7.1. *Monotonicity.* The most important property of the stable norms defined above are their *monotonicity.*

Proposition 1.18 (Monotonicity)**.** *Let $g : X \to X'$ be a map, and let $\alpha \in H_2(X)$. Then $\|g_*\alpha\| \le \|\alpha\|$.*

Similarly, let $g : (X,Y) \to (X',Y')$ be a map of pairs, and let $\alpha \in H_2(X,Y)$. Then $\|g_*\alpha\| \le \|\alpha\|$. Likewise, for any $\gamma \in H_1(Y)$ in the kernel of $H_1(Y) \to H_1(X)$ we have $\text{scl}(g_*\gamma) \le \text{scl}(\gamma)$.

The case of equality is again important enough for special mention:

Definition 1.19. A map g for which equality holds $\|g_*\alpha\| = \|\alpha\|$ for some α, or $\text{scl}(g_*\gamma) = \text{scl}(\gamma)$ for some γ, is said to be *isometric for α or γ*. A map for which equality holds for all α or all γ is said to be *isometric.*

Example 1.20 (Retraction). Suppose $X \subset X'$ (resp. $(X,Y) \subset (X',Y')$) and there is a retraction $r : X' \to X$ (resp. $r : (X',Y') \to (X,Y)$). Then the inclusion of X into X' (resp. (X,Y) into (X',Y')) is isometric.

1.7.2. *Finite covers and transfer.* The various norms behave in a somewhat complicated way under finite covers.

Example 1.21. Let $\pi : \hat{X} \to X$ be a finite cover of degree n. If Y is a subspace of X, let $\hat{Y}$ denote the preimage of Y in $\hat{X}$, so that the restriction $\pi : \hat{Y} \to Y$ is also a finite cover of the same degree. The map π is *not* an isometry in general. The simplest example is when X is a closed, connected nonorientable hyperbolike surface, and $\hat{X}$ is an orientable double cover. Then the fundamental class of $\hat{X}$ has nonzero stable genus (see Example 1.23), but its image has zero stable genus.

On the other hand, there are *transfer maps* $\tau : H_2(X) \to H_2(\hat{X})$, $\tau : H_2(X,Y) \to H_2(\hat{X},\hat{Y})$ and $\tau : H_1(Y) \to H_1(\hat{Y})$. In singular homology these maps are induced at the level of chains, where the transfer of $\sigma : \Delta \to X$ is the formal sum of the (finitely many) lifts $\hat{\sigma} : \Delta \to \hat{X}$. If $f : (S,\partial S) \to (X,Y)$ there is an induced transfer $\tau_* f : (\hat{S},\partial\hat{S}) \to (\hat{X},\hat{Y})$ for a suitable finite cover $\hat{S}$ of S obtained by triangulating S and lifting the triangles one by one. Transfer acts predictably on the various norms:

Proposition 1.22. *Let $\pi : (\hat{X},\hat{Y}) \to (X,Y)$ be a finite cover of degree n. Then $\|\tau_*\alpha\| = n \cdot \|\alpha\|$ and $\mathrm{scl}(\tau_*\gamma) = n \cdot \mathrm{scl}(\gamma)$ for any α in $H_2(X)$ or $H_2(X,Y)$, and any $\gamma \in H_1(Y)$. In particular, the transfer of an extremal surface is extremal for the transfer of the class it represents.*

Proof. The composition of τ with the covering projection π replaces any surface S mapping to X with a finite cover $\hat{S}$ of covering degree n. The result now follows from monotonicity, i.e. from Proposition 1.18 applied to the map π. ☐

1.7.3. *Surfaces.*

Example 1.23 (Surface). Let S be a compact, oriented surface. The identity map $S \to S$ exhibits S as an extremal surface for the fundamental class $[S]$ in $H_2(S)$ or $H_2(S,\partial S)$. Hence $\|[S]\| = -\chi^-(S)/2$, and if S has boundary, $\mathrm{scl}(\partial S) = -\chi^-(S)/2$. It follows from Lemma 1.10 that any covering map $\hat{S} \to S$ is also extremal for $[S]$ and for $[\partial S]$.

On the other hand, if S' is another compact, oriented surface, and $f : (S',\partial S') \to (S,\partial S)$ is not homotopic (as a map of pairs) to an oriented covering map, then S' is *not* extremal.

We will prove these claims in the next section, but for the moment let's just observe that it is just as well they are true, for otherwise the norms $\|\cdot\|$ and scl would be always trivial.

Let's imagine (counterfactually) that the identity map $(S,\partial S) \to (S,\partial S)$ is not extremal for some S. Then S must be hyperbolike (since otherwise $-\chi^-(S)$ is already zero) and there must be some S' and $f : (S',\partial S') \to (S,\partial S)$ a proper map of degree n with $-\chi^-(S')/n = -\chi^-(S) \cdot \lambda$ for some $\lambda < 1$. If $-\chi^-(S') = 0$ then $\|[S]\| = 0$. Otherwise, S' is hyperbolike, and S, S' have common finite covers. But then there is some $S'' \to S'$ proper of degree n' so that $-\chi^-(S'')/n' = -\chi^-(S') \cdot \lambda$ and composing this with f we get a proper map from S'' to S of degree $n \cdot n'$ with $-\chi^-(S'')/(n \cdot n') = -\chi^-(S) \cdot \lambda^2$. By induction this implies $\|[S]\| = 0$ and therefore by taking finite covers, we see that $\|[S']\| = 0$ for *every* surface S'. By monotonicity, this implies that $\|\alpha\| = 0$ for every class in every space. So we see that the counterfactual supposition leads to the triviality of the theory. In fact, the theory will turn out to be nontrivial, so the counterfactual is false.

Example 1.24 (Immersed subsurface). Let S be a compact, oriented hyperbolike surface, and let $\Gamma : L \to S$ be an essential map from an oriented 1-manifold to S. Suppose $\Gamma_*[L] = 0 \in H_1(S)$, so that $\mathrm{scl}(\Gamma)$ is defined. If we choose a hyperbolic structure on S, then after replacing Γ with a homotopic map if necessary, we can assume that $\Gamma(L)$ is a union of immersed geodesics in S. If S' is a surface mapping $f : (S',\partial S') \to (S,\Gamma(L))$ by an *immersion* to S in such a way that $f : \partial S' \to \Gamma(L)$ factors as $\Gamma \circ h$ for some oriented homeomorphism $h : \partial S' \to L$ then S' is extremal for $f_*[S']$; i.e. $\|f_*[S']\| = -\chi^-(S')/2$.

Moreover, *if each component of S has boundary* then S' is extremal for $f_*[\partial S']$; i.e. $\text{scl}(\Gamma) = -\chi^-(S')/2$.

To see this observe that an immersion $S' \to S$ with geodesic boundary lifts to an essential embedding $S' \to \hat{S}$ in some finite cover of S, by Theorem 1.1. The covering map $\hat{S} \to S$ is extremal, by Example 1.23, and therefore so is every essential subsurface, by Lemma 1.11.

After developing some more tools in § 2 we will be able to treat many more examples.

2. Duality

In this section we discuss the duality between stable genus/stable commutator length and two-dimensional bounded cohomology and the theory of *quasimorphisms*. Stable genus and scl are essentially L^1 theories, and the dual theory is an L^∞ theory. Manifestations of L^1–L^∞ duality in special cases gives insight into the geometric meaning of these theories.

2.1. **Min-cut max-flow.** Let X be a directed finite graph (i.e. an ordinary graph in which we have chosen an orientation on each edge). An edge incoming to a vertex v is indicated by $\xrightarrow{e} v$ and an edge outgoing from v is indicated by $v \xrightarrow{e}$.

A *flow* is an assignment of numbers $f(e)$ to each edge (the *flow weight*) satisfying the following two conditions:

(1) For each vertex v,
$$\sum_{\xrightarrow{e} v} f(e) = \sum_{v \xrightarrow{e}} f(e)$$

(2) The flow on each edge does not exceed the capacity: $|f(e)| \le 1$.

The first condition says merely that a flow defines a simplicial 1-cycle on X, so there is a map from the set of flows $\mathcal{F}$ to $H_1(X;\mathbb{R})$. Since X is a graph there are no 2-boundaries, so this map is injective, and we can identify $\mathcal{F}$ with a *subset* of $H_1(X;\mathbb{R})$. The condition that the flow does not exceed the capacity on a given edge imposes a linear inequality, which defines a supporting hyperplane for the set $\mathcal{F}$. In particular, since $\mathcal{F}$ is defined precisely by these inequalities, it is a convex symmetric polyhedron in $H_1(X;\mathbb{R})$. One can therefore think of $\mathcal{F}$ as the unit ball in a *norm* on $H_1(X;\mathbb{R})$. Evidently, this is just the restriction of the unit ball in the L^∞ norm on 1-chains.

A *cut* is a second kind of assignment of numbers $c(e)$ to each edge. A cut determines an element of $H^1(X;\mathbb{R})$ as follows. If every $c(e)$ is an integer, there is a simplicial map $X \to S^1$ sending every vertex to a specific point in S^1, and mapping every (oriented) edge e over S^1 with degree $c(e)$. A homotopy class of map from X to S^1 determines an element of $H^1(X;\mathbb{Z})$, so there is a map from integral cuts to $H^1(X;\mathbb{Z})$. This extends by linearity to a map $\pi : \mathcal{C} \to H^1(X;\mathbb{R})$ where $\mathcal{C}$ is the space of cuts. There is an L^1 norm on $\mathcal{C}$, defined by $\|c\|_1 := \sum |c(e)|$.

There is a natural pairing $\langle \cdot, \cdot \rangle$ between flows and cuts, which is the restriction of the pairing between H_1 and H^1. As a formula, $\langle f, c\rangle = \sum_e f(e)c(e)$. A class $\phi \in H^1$ determines a linear function on H_1, and there is a unique pair of supporting hyperplanes for $\mathcal{F}$ which are level sets of this function. An $f \in \mathcal{F}$ contained in one of these supporting hyperplanes is said to be a *maximal flow*. Similarly, ϕ determines an affine subspace of $\mathcal{C}$, namely the inverse image of ϕ under $\mathcal{C} \to H^1(X)$. A $c \in C$ minimizing the L^1 norm on this subspace is said to be a *minimal cut*.

In this context, the *min-cut max-flow Theorem* is just the observation that for a minimal cut c and a maximal flow f, we have $\|c\|_1 = \langle f, c \rangle$, which is just a restatement of L^1–L^∞ duality (or equivalently, primal-dual duality in linear programming).

2.1.1. *Taut foliations.* Let M be a 3-manifold. A *foliation* $\mathcal{F}$ is a way of filling up M with 2-dimensional surfaces (the *leaves* of the foliation) which locally have the structure of a product. A foliation is *taut* if there is a map from S^1 to M which is transverse to the foliation and intersects every leaf. See [2] for an introduction to the theory of taut foliations.

Let $\mathcal{F}$ be a taut foliation, and fix a triangulation of M so that the leaves of $\mathcal{F}$ are *normal surfaces*; i.e. their intersection with each simplex look like a collection of level sets of a linear function. The 1-skeleton of the triangulation is a graph X, and it may be directed by orienting edges compatibly with a co-orientation on $\mathcal{F}$. The tautness of $\mathcal{F}$ means that we can find such a triangulation and a flow f whose coefficients are all *positive* with respect to the given orientation of edges. The class of f in $H_1(X;\mathbb{R})$ determines a class in $H_1(M;\mathbb{R})$, Poincaré dual to a class in $H^2(M;\mathbb{R})$, represented by a 2-cocycle ω which measures the flux of f locally through a surface. We can think of the absolute value of this flux as defining a combinatorial notion of *area* on normal surfaces in M, and by construction the leaves of the foliation are *locally least area* in their relative homology class. This is essentially just min-cut max-flow again: a normal surface (at least locally) defines a cut by counting the number of intersections with each edge.

This construction makes sense also in the smooth category. The tangent planes to the foliation $\mathcal{F}$ can be defined as the kernel of a nowhere vanishing 1-form α; Frobenius' Theorem says that integrability of $\ker(\alpha)$ is equivalent to the identity $\alpha \wedge d\alpha = 0$. Tautness of $\mathcal{F}$ is equivalent to the existence of a closed 2-form ω on M which restricts to an area form on each leaf of the foliation. There is a unique vector field Y on M tangent to $\ker(\omega)$ and satisfying $\alpha(Y) = 1$ identically; we can think of Y as defining a flow transverse to $\mathcal{F}$. Then $\omega \wedge \alpha$ is a nondegenerate 3-form, and

$$\mathcal{L}_Y(\omega \wedge \alpha) = i_Y d(\omega \wedge \alpha) + d i_Y(\omega \wedge \alpha) = d\omega = 0$$

i.e. the flow Y preserves the form $\omega \wedge \alpha$. We can define a metric g on M for which Y is perpendicular to $\mathcal{F}$, and for which the area form on $\mathcal{F}$ agrees with ω. The form ω becomes a *calibration* of $T\mathcal{F}$ in this metric — i.e. a closed form satisfying $|\omega(\xi)| \le \text{area}(\xi)$ pointwise on 2-planes ξ, and $|\omega(\xi)| = \text{area}(\xi)$ when ξ is tangent to $\mathcal{F}$. In such a metric, leaves of $\mathcal{F}$ are *locally least area*. In fact, if λ is any leaf of $\mathcal{F}$, if $D \subset \lambda$ is a compact subsurface, and if D' is any other surface with $\partial D = \partial D'$, then

$$\text{area}(D) = \int_D \omega = \int_{D'} \omega \le \text{area}(D')$$

In this sense the leaves of $\mathcal{F}$ are "minimal cuts" dual to the "maximal flow" Y.

There is a canonical 2-dimensional class $e_\mathcal{F} \in H^2(M;\mathbb{Z})$ associated to $\mathcal{F}$, called the *Euler class*, which is the obstruction to finding a non-vanishing section of the 2-plane bundle $T^*\mathcal{F}$. For every closed leaf S we have $e_\mathcal{F}([S]) = -\chi(S)$. If S' is another surface in general position with respect to $\mathcal{F}$, there is a section β of $T^*\mathcal{F}|_{S'}$ defined by $\ker(\beta) = TS' \cap T\mathcal{F}$ pointwise. The zeros of β are exactly the points of tangency of S' with $\mathcal{F}$, and for a nondegenerate tangency the local contribution to $e_{\mathcal{F}}([S])$ is ± 1 depending on the parity of the *index* of the singularity (which is even for a local minimum/maximum and odd for a saddle tangency) and whether the orientations of $T\mathcal{F}$ and TS' agree or disagree at the given point.

Now, we have shown how to construct a metric in which leaves of $\mathcal{F}$ are minimal surfaces. Schoen-Yau [10] showed that any map from a surface S' to a Riemannian manifold can either be *compressed* (thereby reducing its complexity) or homotoped to a minimal representative. For such an S' there can be no local minima/maxima of its intersection with $\mathcal{F}$, by the "barrier" property of minimal surfaces, so all the intersections must be saddles. At a saddle intersection the contribution to $e_{\mathcal{F}}([S'])$ is 1 if the orientations of TS' and $T\mathcal{F}$ agree at the point, and -1 otherwise. On the other hand, β is a section of T^*S', and we can compute $-\chi(S')$ by counting the singularities. Now the local contribution depends only on the parity of the index of the singularity. It follows that we derive an inequality

$$e_{\mathcal{F}}([S']) = \sum_{\text{saddles}} \pm 1 \le \sum_{\text{saddles}} 1 = -\chi(S')$$

This argument and the following corollary are due to Thurston [11]:

Proposition 2.1 (Thurston)**.** *Let $\mathcal{F}$ be a taut foliation on a 3-manifold M. Then every closed leaf S is extremal in the homology class $[S]$. That is, $\|[S]\| = -\chi^-(S)/2$.*

For manifolds with boundary and taut foliations transverse to the boundary a similar result holds, as may be seen by doubling.

Conversely, Gabai [5] showed that for every 3-manifold M with $\pi_2(M) = 0$ and every nonzero integral homology class $\alpha \in H_2(M, \partial M; \mathbb{Z})$ there is a taut foliation $\mathcal{F}$ on M with a finite set of closed leaves S representing α. Together with Proposition 2.1 this implies that $\|\alpha\| \in \frac{1}{2}\mathbb{Z}$. A (pseudo)-norm on a finite dimensional vector space taking integer values on an integer lattice has a unit ball which is a finite sided polyhedron (possibly noncompact if the pseudo-norm is degenerate) with rational vertices. We therefore have the following theorem:

Theorem 2.2 (Gabai)**.** *Let M be a 3-manifold with $\pi_2(M) = 0$, and let $\alpha \in H_2(M, \partial M; \mathbb{Z})$. Then $\|\alpha\| \in \frac{1}{2}\mathbb{Z}$ and α is represented by an* embedded *extremal surface. Consequently the unit ball in the norm $\|\cdot\|$ is a finite sided rational polyhedron.*

For a 3-manifold, the quantity $2\|\cdot\|$ is called the *Thurston norm* on $H_2(M, \partial M; \mathbb{Z})$, and is extensively studied by 3-manifold topologists. When M is atoroidal, $\|\cdot\|$ is a genuine norm, and every extremal surface is hyperbolike.

2.2. **Bounded forms.** Let M be a non-positively curved Riemannian manifold, and let ω be a closed 2-form on M, and suppose there is a least (non-negative) constant D_K so that

$$|\omega(\xi)| \le -K(\xi) \cdot D_K/\pi$$

on any oriented 2-plane ξ, where K denotes curvature. Note that such a D_K will always exist when M is compact and strictly negatively curved, and in fact one can think of D_K as a kind of scaled L^∞ (operator) norm on 2-forms. Let L be a totally geodesic submanifold of M on which ω restricts to zero, and let $f : (S, \partial S) \to (M, L)$ be a surface. The form ω defines an element $[\omega]$ of $H^2(M, L)$ which can be paired with S.

There are several different ways to replace the map f by a homotopic map in such a way that the geometry of the image $f(S)$ can be compared with the geometry of M. Call a map *efficient* if it satisfies the following two properties:

(1) for each $p \in S$ the intrinsic curvature K_S of $f(S)$ at $f(p)$ is bounded above by the curvature of M along the 2-plane $df(T_pS)$; and
(2) the geodesic curvature k_g of $f(S)$ along ∂S is non-positive.

For an efficient map, the Gauss-Bonnet theorem gives

$$[\omega]([S])/4D_K = \int_S \omega/4D_K \le \int_S -K_S/4\pi \le -\left(\int_S K_S + \int_{\partial S} k_g\right)/4\pi = -\chi(S)/2$$

so the value of $[\omega]([S])/4D_K$ gives a lower bound for the stable (relative) genus. Moreover, a surface S for which $f(S)$ is totally geodesic and satisfies $\omega(\xi) = -K(\xi) \cdot D_K/\pi$ pointwise for ξ in f_*TS will be extremal. This is a very effective method to produce examples of extremal surfaces, especially in spaces M with a lot of symmetry.

There are many natural ways to produce efficient representatives in a relative homotopy class when the target is non-positively curved, including the following:

(1) (harmonic) Choose a conformal structure on S and let f be the unique harmonic map in its homotopy class. Eells-Sampson [4] showed that such a map exists whenever M is non-positively curved. For S with boundary, one must first fix f on the boundary by taking ∂S to a geodesic representative in L (which will also be geodesic in M since L is totally geodesic).
(2) (minimal surfaces) As discussed in § 2.1, Schoen-Yau [10] show in great generality that for Riemannian targets M with $\pi_2(M) = 0$ (which holds automatically when M is complete and non-positively curved) every (relative) map from a surface S to M can either be *compressed* (thereby reducing $-\chi^-(S)$) or homotoped to a locally least area minimal surface. For S with boundary we must also allow the possibility that S can be *boundary compressed* — i.e. that there is a proper essential arc β in S whose endpoints both map to the same point, and for which $f(\beta)$ is a contractible loop in X.
(3) (polygonal and pleated surfaces) Choose an embedded essential loop γ in S (for example, γ could be a component of ∂S if S has boundary) and a triangulation with all vertices on $\partial S \cup \gamma$. We can build a map $f : S \to M$ by first making f geodesic on $\partial S \cup \gamma$, and then edge by edge on the 1-skeleton of the triangulation. Then fill in f over each triangle with a ruled surface — i.e. a union of geodesics from one vertex to the points on the opposite side. The resulting surface is efficient, and might additionally have atoms of negative curvature at the vertices. When M is strictly negatively curved, one can obtain a limit of such surfaces obtained by *spinning* each vertex around γ or the component of ∂S on which it lies; the result is a *pleated surface* made up of *ideal triangles* in which the vertices have been dragged off to infinity. See [12] § 8 for an introduction to the theory of pleated surfaces.

It follows that if M is closed and strictly negatively curved, $\|\alpha\| > 0$ for any class α in $H_2(M; \mathbb{R})$.

2.2.1. *Estimates from triangles and from pants.* In fact, it is not necessary to obtain a pointwise comparison between ω and the curvature of M to get lower bounds on stable genus. A surface S of genus g with b boundary components may be triangulated with $4g+3b-4$ triangles, but if we first take a finite cover $\hat{S}$ of degree n and the same number of boundary components, then we need only $-2n\chi^-(S)+O(1)$ triangles to triangulate $\hat{S}$. Since M is non-positively curved, $\pi_2(M)$ is trivial, so for any triangle Δ the integral $\int_\Delta \omega$ depends only on $\partial\Delta$. Hence the details of how we fill in a map defined on the 1-skeleton of a triangulation are irrelevant.

Suppose we know that there is a least constant D_Δ so that $|\int_\Delta \omega| \le D_\Delta$ for any geodesic triangle Δ. Then

$$[\omega]([S]) = \frac{1}{n}[\omega]([\hat{S}]) \le -2D_\Delta \chi^-(S) + O(1/n)$$

and therefore taking the limit as $n \to \infty$ we get the inequality $\|[S]\| \ge [\omega]([S])/4D_\Delta$.

Instead of using triangles, we can use pants. Now there is no need to pass to covers, since any hyperbolike surface S can be decomposed into $-\chi(S)$ pants. Suppose there is a least constant D_P so that $|\int_P \omega| \le D_P$ for any pants P in M with geodesic boundary. Then $[\omega]([S]) \le -D_P\chi^-(S)$, and therefore $\|[S]\| \ge [\omega]([S])/2D_P$.

Now, every pants with geodesic boundary can be decomposed into 2 ideal triangles, and therefore $D_P \le 2D_\Delta$, so the second inequality is at least as good as the first inequality.

If M is closed and strictly negatively curved, then intermediate between geodesic triangles and geodesic pants we can consider triangulations by *ideal* triangles. Again, we can let $D_{\mathring{\Delta}}$ be the least constant so that $|\int_{\mathring{\Delta}} \omega| \le D_{\mathring{\Delta}}$ for all ideal triangles. Then $\|[S]\| \ge [\omega]([S])/4D_{\mathring{\Delta}}$ and we have the inequality $D_P \le 2D_{\mathring{\Delta}} \le 2D_\Delta$ for any ω.

On the other hand, if M is closed of strictly negative curvature, the first inequality is actually an *equality*:

Proposition 2.3. *Let M be closed of strictly negative curvature, and let ω be a closed 2-form. Then $D_P = 2D_{\mathring{\Delta}}$.*

Proof. By approximating ideal triangles by ordinary triangles, for any small ϵ and large T we can find a geodesic triangle Δ with all edges of length $\ge T$ and for which $\int_\Delta \omega \ge D_{\mathring{\Delta}} - \epsilon$. By the ergodicity of the geodesic flow, for each vertex p of Δ and each unit vector $v \in T_pM$ there is a geodesic γ emanating from p with the angle between γ' at its initial point and v as small as we like so that γ eventually returns to p, and the angle between γ' at its endpoint and $-v$ is also as small as we like. Since Δ has all edges of length $\ge T$, the angles at the vertices are all as small as we like. So we can find three geodesic segments as above each of which starts and ends at a given vertex, and makes an arbitrarily small angle with both edges of Δ at both endpoints. From two copies of $\partial\Delta$ and two copies of each of these three geodesic segments we can make three piecewise geodesic loops which form the boundary of a degenerate pair of pants P' whose interior is made from two copies of Δ. This can be straightened to a genuine pair of pants P in such a way that P and P' are ϵ-close (in C^∞) on $1-\epsilon$ of their area. Thus

$$\left| \int_P \omega - 2\int_\Delta \omega \right| \le O(\epsilon)$$

Taking $\epsilon \to 0$ completes the proof. ∎

2.3. **1-forms.** Let θ be a closed 1-form on L pairing nontrivially with some $\gamma \in H_1(L)$ in the kernel of $H_1(L) \to H_1(M)$. Since this class is trivial in $H_1(M)$, we cannot extend θ to a closed 1-form on M, but we can try to extend it to a form (which, by abuse of notation, we denote θ) for which $\omega := d\theta$ is as small as possible, as measured (for example) by D_K, D_Δ, $D_{\mathring{\Delta}}$ or D_P.

Let's continue to assume that M is closed of strictly negative curvature. If $g \in \pi_1(M)$ then the conjugacy class of g is represented by a unique closed oriented geodesic γ_g, and we can define a class function $\alpha_\theta : \pi_1(M) \to \mathbb{R}$ by

$$\alpha_\theta(g) = \int_{\gamma_g} \theta$$

For any g, h the three loops $\gamma_{gh}, \gamma_{g^{-1}}, \gamma_{h^{-1}}$ form the three oriented boundary components of a pair of pants P. By Stokes' theorem,

$$|\alpha_\theta(gh) - \alpha_\theta(g) - \alpha_\theta(h)| = \left| \int_{\partial P} \theta \right| = \left| \int_P \omega \right| \le D_P$$

Since any surface can be decomposed into pairs of pants, we obtain an inequality

$$\text{scl}(\gamma) \ge \alpha_\theta(\gamma)/2D_P$$

2.3.1. *Antisymmetrizing metrics.* It is a bit irritating to define the function $\alpha_\theta(g)$ in terms of the geodesic γ_g, which in turn depends on an external parameter, namely the Riemannian metric. It is more elegant to combine the two parameters, as follows. For a tangent vector v, let $\|v\|$ denote its length; then for each real number t we can define $\|v\|_t = \|v\| + t\theta(v)$. For $t \cdot \|\theta\|_\infty < 1$ this defines an *asymmetric* norm on the tangent space at each point, and for any two points p, q we can define

$$d_t(p, q) = \inf_\gamma \int_0^1 \|\gamma'(s)\|_t ds$$

where the infimum is taken over all paths $\gamma : [0, 1] \to M$ with $\gamma(0) = p$ and $\gamma(1) = q$. The function d_t is an asymmetric metric, i.e. it satisfies all the axioms of a (path) metric except for symmetry. For any conjugacy class $g \in \pi_1(M)$ we can define $\ell_t(g)$ to be the infimum of the (oriented) length of an oriented loop in the free homotopy class of g. Then we can recover α_θ by *antisymmetrizing* d_t; that is, we can define $\alpha_{\theta,t}$ by

$$\alpha_{\theta,t}(g) = \frac{1}{2t}\left(\ell_t(g) - \ell_t(g^{-1})\right)$$

and then observe that $\alpha_\theta = \lim_{t \to 0} \alpha_{\theta,t}$.

Having taken this step, it is natural to generalize to an arbitrary antisymmetric metric d on M, with associated length function ℓ, and define $\alpha(g) = \ell(g) - \ell(g^{-1})$ and $D_P = \sup_{g,h} |\alpha(gh) - \alpha(g) - \alpha(h)|$. If M (closed) admits *some* metric of negative curvature, then D_P is finite, and we get a nontrivial estimate of the form $\text{scl}(\gamma) \ge \alpha(\gamma)/2D_P$.

2.4. **Causal structures.** A *cone* in a real vector space is a closed, strictly convex subset invariant under multiplication by a positive real number. A *cone field* C on a smooth oriented manifold M is a choice of cone in the tangent space at each point, varying continuously on compact subsets. For a point p of M, let C_p denote the cone of C in T_pM. A vector v in T_pM is *positive (resp. negative) timelike* if it is contained in the interior of

C_p (resp. the interior of $-C_p$), is *positive (resp. negative) lightlike* if it is contained in the frontier of C_p (resp. the frontier of $-C_p$) and is *spacelike* if it is contained in the complement of $C_p \cup -C_p$. The standard example of a cone field is that arising from a Lorentz metric with a time orientation.

If M is a manifold with a cone field C, we say that a map $\gamma : L \to M$ from a 1-manifold to M is *timelike* if γ' is contained in C at each point (and we define lightlike and spacelike 1-manifolds similarly).

Example 2.4. The oriented circle S^1 admits a cone field where, for each p, the positive cone C_p consists of vectors tangent to p oriented positively.

Example 2.5 (Circle bundle). If M is a smooth manifold, a closed 2-form ω is the curvature of an S^1 bundle E with a unitary connection if and only if it has periods in $2\pi\mathbb{Z}$ on $H_2(M;\mathbb{Z})$. If ω is exact, so that $\omega = d\theta$ for some 1-form, the bundle is trivial, and we may take θ to be the connection 1-form relative to some (global) trivialization. The horizontal distribution ξ (associated to the connection) determines a splitting of TE as $V \oplus \xi$ where V is the (real) line bundle tangent to the circle fibers. We can define a Lorentz metric h on E by $h = -dt^2 \oplus \pi^* g$ where $\pi : E \to M$ is projection to M, and g is a Riemannian metric on M. If $|\omega| \le -K$ then every closed timelike curve is homotopically essential.

A *Cauchy hypersurface* is a proper, oriented spacelike submanifold $Y \subset M$ of codimension 1 which intersects every closed timelike curve. If M is n-dimensional, then $(Y, \partial Y)$ determines a class $[Y] \in H_{n-1}(M, \partial M; \mathbb{Z})$ Lefschetz dual to some $\theta_Y \in H^1(Ml\mathbb{Z})$. If $\gamma : L \to M$ is a timelike (oriented) 1-manifold, the *winding number* of γ is $\theta_Y(\gamma_*[L])$. Equivalently, it is the number of intersections of L with Y (note that all these intersections will have positive sign). By abuse of notation, we abbreviate $\theta_Y(\gamma_*[L])$ by $\text{wind}(\gamma)$. Note by hypothesis that $\text{wind}(\gamma)$ is strictly positive for every (nonempty) γ.

A cone field C is *recurrent* if there is a (nontrivial) timelike curve from any point p to any other point q. Let's suppose that C is recurrent, that M is compact, and there is a (closed) Cauchy hypersurface Y. Since C is recurrent, every two points p, q are contained on a closed timelike loop, and by compactness there is some least (positive) integer w so that every two points are contained on a timelike loop γ with $\text{wind}(\gamma) \le w$.

Let $\hat{M}$ denote the cyclic cover of M dual to θ_Y, and $\hat{C}$ the lifted cone field on $\hat{M}$. Let G be a group of diffeomorphisms of M preserving the cone field C, and let $\hat{G}$ be the central extension of G consisting of lifts of elements of G to $\hat{M}$, so that $\hat{G}$ preserves the cone structure $\hat{C}$. The deck group of the cover $\hat{M}$ is $\mathbb{Z}$, and we write the action of this deck group on $\hat{M}$ by $p \to p + n$ for $n \in \mathbb{Z}$. Note that the deck group is a central subgroup of $\hat{G}$.

We define a relation $\prec$ on $\hat{M}$ by $p \prec q$ if there is a timelike curve from p to q. Note that this relation is transitive, and defines a partial order on $\hat{M}$. For any two points $p, q \in \hat{M}$ we define $d(p, q)$ to be the greatest integer n so that $p \prec q - n$. This is well-defined, since we assumed that C was recurrent. Then we can define a *rotation number* $\rho : \hat{G} \to \mathbb{R}$ by

$$\rho(g) = \lim_{n \to \infty} \frac{d(p, g^n(p))}{n}$$

and observe that this does not depend on the choice of $p \in \hat{M}$.

Lemma 2.6. *For any commutator $[\alpha,\beta]$ in $\hat{G}$ and for any $q \in \hat{M}$ there is an estimate*

$$q - 2w \prec [\alpha,\beta](q) \prec q + 2w$$

Proof. First, observe that the value of $[\alpha,\beta]$ is unchanged if we multiply α or β by any element of the center. So for any $p \in \hat{M}$ and any two elements $\alpha,\beta \in \hat{G}$ we can adjust α and β by an element of the center so that $p \preceq \alpha(p) \prec p + w$ and similarly for β. Hence

$$p \preceq \alpha(p) \preceq \alpha\beta(p) \prec \alpha(p+w) \prec p + 2w$$
$$p \preceq \beta(p) \preceq \beta\alpha(p) \prec \beta(p+w) \prec p + 2w$$

Set $q = \beta\alpha(p)$. Then $p \preceq q \prec p + 2w$ and we deduce

$$q - 2w \prec p \preceq \alpha\beta(p) = [\alpha,\beta](q) \prec p + 2w \preceq q + 2w$$

and the lemma is proved. □

It follows immediately that $\mathrm{scl}(g) \le |\rho(g)|/2w$ for every $g \in \hat{G}$.

2.5. **Examples.**

Example 2.7. The *hyperbolic plane* $\mathbb{H}^2$ is the unique complete Riemannian 2-manifold of constant curvature -1. Because it is 2-dimensional, the area form ω is closed, and satisfies $\omega(\xi) = -K(\xi)$ pointwise on 2-planes whose orientation agrees with that of $\mathbb{H}^2$. If Γ is a discrete, torsion-free group of isometries of $\mathbb{H}^2$, the quotient $M = \mathbb{H}^2/\Gamma$ is a 2-dimensional hyperbolic surface. Any homotopically essential 1-manifold in M is homotopic to a union of geodesics L. Any positively oriented immersion $f : (S, \partial S) \to (M, L)$ is therefore extremal, justifying the claims made in Example 1.23 and Example 1.24. There is a hyperbolic space $\mathbb{H}^n$ of any dimension n, but only in dimension 2 does it come with a *natural* symplectic form (i.e. one invariant under the full orientation-preserving isometry group).

Example 2.8. *Complex hyperbolic space* of dimension n, denoted $\mathbb{CH}^n$, is the symmetric space obtained as the quotient $U(n,1)/U(n) \times U(1)$ where $U(n,1)$ is the space of complex $(n+1)\times(n+1)$ matrices preserving the form $\langle w, z\rangle = w_1\bar{z}_1 + \cdots w_n\bar{z}_n - w_{n+1}\bar{z}_{n+1}$ on $\mathbb{C}^{n+1}$ In fact, $U(n,1)$ acts transitively on the "hyperboloid"

$$H := \{w \in \mathbb{C}^{n+1} \text{ with } \langle w, w\rangle = -1\}$$

with point stabilizers the conjugates of $U(n)$, and $\mathbb{CH}^n$ is obtained as the quotient of H by the action of $U(1)$ acting diagonally on w by multiplying its coordinates by $e^{i\theta}$. This exhibits $\mathbb{CH}^n$ as an open domain in complex projective space, and gives it the structure of a complex manifold.

The inner product $\langle\cdot,\cdot\rangle$ restricts to a Lorentz metric on H of (real) signature $(2n, 1)$, and the transitivity of the action reveals H to be a space of constant Lorentz curvature -1. The image of the real hyperboloid $H_{\mathbb{R}} := H \cap \mathbb{R}^{n+1}$ in $\mathbb{CH}^n$ is a totally geodesic totally real copy of ordinary hyperbolic space $\mathbb{H}^n$. On the other hand, the exceptional isomorphism $\mathfrak{su}(1,1) = \mathfrak{o}_{\mathbb{R}}(2,1)$ (an unusual real form of the more familiar $\mathfrak{su}(2) = \mathfrak{o}_{\mathbb{R}}(3)$) shows that the translates of $\mathbb{CH}^1$ in $\mathbb{CH}^n$ are isometric to *scaled* copies of $\mathbb{H}^2$. O'Neill's formula, together with the fact that H has constant (Lorentz) curvature -1, shows that the (Riemannian) curvature of $\mathbb{CH}^n$ varies between -1 on the translates of $\mathbb{H}^n$, and -4 on the translates of $\mathbb{CH}^1$.

The metric and complex structure on $\mathbb{CH}^n$ are compatible, and give it the natural structure of a Kähler manifold. The symplectic form ω measures the curvature of H as a circle bundle over $\mathbb{CH}^n$. Consequently, ω vanishes on $\mathbb{H}^n$ (i.e. these are Lagrangian submanifolds), and is maximal on the copies of $\mathbb{CH}^1$. Now, suppose Γ is a discrete subgroup of $U(n,1)$, and $M = \mathbb{CH}^n/\Gamma$ is the quotient manifold. A map $f : S \to M$ where S is a closed, connected, oriented surface, determines a representation $\pi_1(S) \to U(n,1)$. If the image stabilizes a totally geodesic $\mathbb{CH}^1$ in $\mathbb{CH}^n$, the quotient is a totally geodesic holomorphic surface in M, and by the argument above it is extremal for the class $f_*[S] \in H_2(M;\mathbb{R})$.

Example 2.9. The *Siegel upper half-space* $\mathfrak{H}_g$ is the symmetric space obtained as the quotient $\mathrm{Sp}(2n,\mathbb{R})/U(n)$ where we identify $\mathbb{C}^n$ with the standard symplectic $\mathbb{R}^{2n}$ in order to define the inclusion of $U(n)$ into $\mathrm{Sp}(2n,\mathbb{R})$. There is a $\mathbb{C}^n$ bundle over $\mathfrak{H}_g$, obtained by thinking of $\mathrm{Sp}(2n,\mathbb{R}) \to \mathfrak{H}_g$ as a principal $U(n)$ bundle, and the associated determinant line bundle has a curvature form ω that defines a natural symplectic structure on $\mathfrak{H}_g$.

2.6. **Quasimorphisms and bounded cohomology.** The examples and constructions from the previous sections can be abstracted. If X is a space, the singular chain groups $C_i(X;\mathbb{R})$ come together with canonical bases, namely the continuous maps of the i-simplex to X. These canonical bases determine canonical L^p norms on $C_i(X;\mathbb{R})$ for every p. Dual to $C_i(X;\mathbb{R})$ with its L^p norm are the i-cocycles of finite L^q norm, where $1/p + 1/q = 1$; in particular, dual to C_i with its L^1 norm are the *bounded* i-cocycles $C_b^i(X;\mathbb{R})$ — those whose absolute value on all i-simplices is bounded by some constant. The coboundary operator is continuous in the L^q norm for any q, and we thereby obtain the complex of *bounded cocycles*, and its associated *bounded cohomology groups* $H_b^*(X;\mathbb{R})$. The L^∞ norm on C_b^* restricts to a norm on the space of bounded cocycles, and determines a quotient pseudo-norm on $H_b^*(X;\mathbb{R})$.

If G is a group, we can similarly form the *bar complex* $C_i(G;\mathbb{R})$ which has a canonical basis consisting of i-tuples $(g_1, g_2, \cdots, g_i) \in G^i$. We can think of the vertices of an i-simplex as the integers from 0 to i, and g_j as a label on the edge from $j-1$ to j. Restricting to codimension one faces (and relabeling vertices in an order-preserving way) determines the boundary maps in the bar complex. In dimension 2 the boundary map is

$$\partial : (g,h) \to g + h - gh$$

Exactly as for spaces we form the bounded cochains and cohomology groups $C_b^i(G;\mathbb{R})$ and $H_b^i(G;\mathbb{R})$. Spelled out, the L^∞ norm of an i-cochain ϕ is defined by

$$\|\phi\|_\infty := \sup \|\phi(g_1, g_2, \cdots, g_i)\|$$

where the supremum is taken over all i-tuples of elements in G.

Proposition 2.10. *Bounded cohomology satisfies the following properties:*

1. *If X is a connected space and $G = \pi_1(X)$, then $H_b^*(X;\mathbb{R})$ and $H_b^*(G;\mathbb{R})$ are canonically isometrically isomorphic.*
2. *$H_b^1(G;\mathbb{R})$ is always trivial. If G is amenable, then $H_b^*(G;\mathbb{R}) = 0$ in every dimension.*
3. *There are comparison maps $H_b^*(G;\mathbb{R}) \to H^*(G;\mathbb{R})$ from bounded to ordinary cohomology. If G is a hyperbolic group, this map is* surjective *for dim > 1.*

Bullets (1) and (2) are proved by Gromov [7]; bullet (3) is due to Mineyev [8].

In low dimensions this specializes as follows. A real (group) 1-cocycle is a homomorphism to $\mathbb{R}$ — in other words an element of $H^1(G;\mathbb{R})$ — and is never bounded unless it is trivial. There is a map from bounded 2-cocycles $Z_b^2(G;\mathbb{R})$ to ordinary cohomology $H^2(G;\mathbb{R})$, and everything in the kernel is $\delta\theta$ for some 1-cochain θ, unique up to a 1-cocycle.

Definition 2.11. Let G be a group. A function $\theta : G \to \mathbb{R}$ is a *quasimorphism* if there is some least number $D(\theta)$ (called the *defect*) such that

$$|\theta(g) + \theta(h) - \theta(gh)| \le D(\theta)$$

for all $g, h \in G$. The set of all quasimorphisms on G is a real vector space, and is denoted $\hat{Q}(G)$.

A 1-cochain is the same thing as a function on G. From the definition of the L^∞ norm and the coboundary operator, a 1-cochain θ has the property that $\delta\theta$ is bounded if and only if θ is a quasimorphism.

Two 1-cochains have the same image in H_b^2 under δ if and only if they differ by a bounded function plus a homomorphism to $\mathbb{R}$. There is therefore an exact sequence

$$0 \to H^1(G;\mathbb{R}) \to \hat{Q}(G)/C_b^1(G;\mathbb{R}) \xrightarrow{\delta} H_b^2(G;\mathbb{R}) \to H^2(G;\mathbb{R})$$

where the last map is not surjective in general. Note that $D(\cdot)$ defines a pseudo-norm on $\hat{Q}$ which vanishes exactly on the subspace H^1, and for a class $\alpha \in H_b^2$ in the kernel of the map to ordinary H^2, the norm $\|\alpha\|_\infty$ is equal to the infimum of $D(\theta)$ over all θ with $[\delta\theta] = \alpha$ (the set of such θ is a coset of C_b^1). Actually, it is straightforward to see that $\hat{Q}/H^1$ is a Banach space in the $D(\cdot)$ norm, and that $D(\theta) = [\delta\theta]$ for some θ in every coset of C_b^1; however such a θ will not in general be unique in its coset.

2.6.1. *Homogeneous quasimorphisms.* It is inconvenient to work with the quotient space $\hat{Q}/C_b^1$, and much more pleasant to choose a canonical representative of each coset.

Definition 2.12. Let G be a group. A quasimorphism θ is *homogeneous* if $\theta(g^n) = n\theta(g)$ for all $g \in G$ and all integers n. The set of all homogeneous quasimorphisms on G is a real vector space, and is denoted $Q(G)$.

Proposition 2.13. *On any group, a homogeneous quasimorphism is a class function, and its restriction to any amenable subgroup is a homomorphism. Moreover, for any (not necessarily homogeneous) quasimorphism $\theta \in \hat{Q}(G)$ the function $\bar{\theta}$ defined by the* homogenization *procedure*

$$\bar{\theta}(g) := \lim_{n\to\infty} \theta(g^n)/n$$

is well-defined, is the unique homogeneous quasimorphism in the coset of C_b^1 containing θ, and satisfies $D(\bar{\theta}) \le 2D(\theta)$.

It follows that we can instead write

$$0 \to H^1(G;\mathbb{R}) \to Q(G) \xrightarrow{\delta} H_b^2(G;\mathbb{R}) \to H^2(G;\mathbb{R})$$

Now we have inequalities $\|[\delta\theta]\|_\infty \le D(\theta) \le 2\|[\delta\theta]\|_\infty$ for any $\theta \in Q$. The following question is fundamental; as far as I know, it was first formulated by the author and Marc Burger following discussions in 2007:

Question 2.14. *Let G be a group and let θ be a homogeneous quasimorphism. Does $D(\theta) = 2\|[\delta\theta]\|_\infty$?*

Proposition 2.3 is the analog of Question 2.14 in the de Rham world, for strictly negatively curved manifolds. Note that to answer Question 2.14 it suffices to answer it for free groups, since any quasimorphism on any group pulls back to a quasimorphism on a free group.

2.6.2. *Harmonic quasimorphisms.* Let G be a finitely generated group, and let μ be a probability measure on G of finite first moment. This means that for a μ-random element $g \in G$, the expectation $\mathbb{E}(|g|) < \infty$, where $|g|$ denotes the word length of g with respect to some finite generating set. The measure μ is *symmetric* if $\mu(g) = \mu(g^{-1})$ for any g, and *nondegenerate* if the support of μ generates G.

A function $f : G \to \mathbb{R}$ is *μ-harmonic* if $\mathbb{E}(f(hg)) = f(h)$ for all h, and for g a μ-random element of G. If we define μ-random walk $\text{id} =: h_0, h_1, h_2, \cdots$ by $h_i = h_{i-1}g_i$ where the g_i are independent μ-random variables, then to say that f is μ-harmonic is to say that the random variables $f_i := f(h_i)$ are a *martingale*.

The following is Cor. 3.14 from [1]:

Theorem 2.15 (Burger-Monod)**.** *Let G be a finitely generated group, and μ a probability measure of finite first moment. Then for every quasimorphism θ there is an antisymmetric μ-harmonic quasimorphism θ_μ in the same C^1_b coset as θ.*

Hüber ([6], unpublished) showed that θ_μ as above is *unique*, and minimizes the defect in its coset:

Proposition 2.16 (Hüber)**.** *Let G and μ be as above, and let θ_μ be an antisymmetric μ-harmonic quasimorphism. Then θ_μ is unique in its C^1_b coset, and achieves the infimum of defect there. In other words, $D(\theta_\mu) = \|[\delta\theta_\mu]\|_\infty$.*

Thus Question 2.14 reduces to the question of whether $D(\theta) = 2D(\theta_\mu)$ for every $\theta \in Q$ and every μ.

2.7. **The duality theorem.**

3. Free groups and surface groups

The fundamental group of a compact connected surface with boundary is free. Thus the theory of stable genus and stable commutator length in free and surface groups has a lot to say about the category of compact, oriented surfaces, and homotopy classes of maps between them. Geometric group theorists have studied the *automorphisms* in this category intensively — this is essentially the theory of mapping class groups, and of $\text{Out}(F)$. But by studying families of maps between surfaces of all topological types simultaneously, different kinds of structures emerge. In this section we discuss some of these structures.

3.1. **The rationality theorem.**

3.2. **The rigidity theorem.**

3.3. **Sails and surgery.**

3.4. Greedy rationals and fractal polyhedra.

4. Symplectic geometry

5. Acknowledgments

Danny Calegari was supported by NSF grant DMS 1005246.

References

[1] M. Burger and N. Monod, *Bounded cohomology of lattices in higher rank Lie groups*, J. Eur. Math. Soc. **1** (1999), no. 2, 199–235
[2] D. Calegari, *Foliations and the geometry of 3-manifolds*, Oxford Mathematical Monographs. Oxford University Press, Oxford, 2007
[3] D. Calegari, *scl*, MSJ Memoirs **20** Mathematical Society of Japan, Tokyo, 2009
[4] J. Eells and J. Sampson, *Harmonic mappings of Riemannian manifolds*, Amer. J. Math. **86** (1964), no. 1, 109–160
[5] D. Gabai, *Foliations and the topology of* 3*-manifolds*, J. Diff. Geom. **18** (1983), no. 3, 445–503
[6] T. Hüber, *A note on harmonic quasimorphisms*, unpublished preprint, May 2010
[7] M. Gromov, *Volume and bounded cohomology*, IHES Publ. Math. **56** (1982), 5–99
[8] I. Mineyev, *Straightening and bounded cohomology of hyperbolic groups*, Geom. Funct. Anal. **11** (2001), no. 4, 807–839
[9] P. Scott, *Subgroups of surface groups are almost geometric*, J. London Math. Soc (2) **17** (1978), no. 3, 555–565
[10] R. Schoen and S.-T. Yau, *Existence of incompressible minimal surfaces and the topology of three dimensional manifolds with non-negative scalar curvature*, Ann. Math. **110** (1979), 127–142
[11] W. Thurston, *A norm for the homology of* 3*-manifolds*, Mem. AMS **59** (1986), no. 339, i–vi and 99–130
[12] W. Thurston, *The geometry and topology of three-manifolds*, a.k.a. *Thurston's notes*, available from the MSRI

University of Chicago, Chicago, Ill 60637 USA
E-mail address: dannyc@math.uchicago.edu